

Maurício Pamplona Segundo

**SEGMENTAÇÃO FACIAL E DETECÇÃO DE CARACTERÍSTICAS
FACIAIS EM IMAGENS DE PROFUNDIDADE**

Curitiba - PR, Brasil

18 de março de 2010

Maurício Pamplona Segundo

**SEGMENTAÇÃO FACIAL E DETECÇÃO DE CARACTERÍSTICAS
FACIAIS EM IMAGENS DE PROFUNDIDADE**

Dissertação apresentada para obtenção do título de
Mestre em Informática pela Universidade Federal do
Paraná.

Orientador: Luciano Silva, Olga Bellon

Curitiba - PR, Brasil

18 de março de 2010

Agradecimentos

O autor gostaria de agradecer ao CNPq, à FINEP e à CAPES pelo suporte financeiro; ao Dr. Jonathon Phillips e ao Dr. Patrick Flynn por permitirem o uso das bases FRGC v1.0 e v2.0; ao Dr. Lijun Yin e *The Research Foundation of State University of New York* por permitirem o uso da base BU-3DFE; e à Grazielle Vernize e ao Paulo Ferreira pela marcação manual das características faciais nas imagens das bases FRGC v1.0 e v2.0.

Sumário

Lista de Figuras	p. iii
Lista de Tabelas	p. vi
Resumo	p. vii
Abstract	p. viii
1 Introdução.....	p. 7
2 Segmentação facial em imagens de profundidade	p. 10
2.1 Localização de regiões homogêneas	p. 11
2.2 Identificação de regiões da face	p. 13
3 Detecção de características faciais	p. 17
3.1 Classificação de curvatura de superfícies	p. 18
3.2 Detecção de características	p. 19
4 Resultados experimentais.....	p. 23
4.1 Resultados da segmentação	p. 24
4.2 Resultados da detecção de característica	p. 27
4.3 Normalização facial	p. 34
5 Conclusão	p. 36
Referências Bibliográficas	p. 38

Lista de Figuras

Figura 2.1	Diagrama da nossa abordagem de segmentação.	10
Figura 2.2	(a) Histograma de gradiente de todas as imagens da base FRGC v2.0. O ponto de inflexão representa o limiar global de gradiente; (b)-(c) histograma de valores de resolução de todas as imagens da base FRGC v2.0 utilizando o eixo y e o eixo x, respectivamente.	12
Figura 2.3	(a) Regiões homogêneas após a rotulação; (b)-(d) mapas de distância destas regiões.	14
Figura 2.4	Detecção de elipse através da combinação de regiões homogêneas.	14
Figura 2.5	(a)-(c) Regiões incluídas na elipse detectada utilizadas para localizar a face; e (d) um exemplo de face dividida por causa do bigode.	16
Figura 3.1	Diagrama da nossa abordagem de detecção de características faciais.	17
Figura 3.2	Exemplo da classificação de superfícies aplicada em imagens de uma mesma pessoa apresentando diferentes expressões: (a), (b) e (c) são os resultados da segmentação; (d), (e) e (f) mostram a classificação de superfícies das imagens (a), (b) e (c), respectivamente. Regiões em vermelho mostram picos na superfície e regiões em amarelo mostram poços.	18
Figura 3.3	Exemplo de detecção da coordenada y da ponta do nariz.	20
Figura 3.4	Exemplo da detecção da coordenada y dos olhos, base do nariz e boca.	20

Figura 3.5	Exemplo da detecção da coordenada x da ponta do nariz.	21
Figura 3.6	Exemplo da detecção dos cantos do nariz.	22
Figura 3.7	Exemplo da detecção dos cantos dos olhos.	22
Figura 4.1	Segmentações incorretas causadas por (a)-(e) regiões irrelevantes selecionadas como face e (f) partes da face não selecionadas.	24
Figura 4.2	Exemplos da segmentação facial utilizando a abordagem proposta em imagens com diferentes dificuldades: (a)-(c) cabelo; (d)-(e) barba e bigode; e (f) expressão facial.	25
Figura 4.3	Exemplo da segmentação facial através de uma técnica padrão [1]: (a) imagem de intensidade, (b) mapa de profundidade, (c) imagem de intensidade após remoção do fundo, e (d) resultado da extração facial.	26
Figura 4.4	Exemplos da segmentação facial obtidas pela técnica padrão de segmentação.	26
Figura 4.5	(a) Ilustração da abordagem proposta por Mian (imagem retirada de [2]), e (b)-(d) resultados obtidos por esta abordagem.	27
Figura 4.6	Comparação entre abordagens de segmentação utilizando (a) ICP+MSAC e (b) SA+SIM.	28
Figura 4.7	Avaliação da detecção de características: LDH para (a) ponta do nariz, (b) cantos internos dos olhos e (c) cantos do nariz.	31
Figura 4.8	Resultados da nossa abordagem de detecção de características faciais.	32

Figura 4.9	Exemplo das técnicas para aprimorar o reconhecimento baseadas em características (imagens retiradas de [3]): (a) área circular em torno do nariz, (b) área elíptica em torno do nariz, (c) parte superior da face, e (d) abordagem com setores priorizados.	32
Figura 4.10	Exemplo da aplicação da normalização facial: (a) imagem original, (b) curva de perfil identificada através do processo de normalização, e (c) imagem normalizada.	34
Figura 4.11	Taxas de verificação obtidas para as imagens antes e depois da normalização.	35

Lista de Tabelas

Tabela 3.1	Classificação dos tipos de superfícies baseada nos valores de curvatura Gaussiana e curvatura média [4].	19
Tabela 4.1	Resultados para a base FRGC v1.0: erro médio e desvio padrão para a nossa abordagem e também para [5, 6]; taxas de detecção com limiar de 10mm (T10) para a nossa abordagem e para [6]; e taxas de detecção com limiar de 15mm (T15) para a nossa abordagem. PN=ponta do nariz; OE=olho esquerdo; OD=olho direito; NE=canto esquerdo do nariz; ND=canto direito do nariz.	29
Tabela 4.2	Resultados para a base FRGC v2.0: erro médio e desvio padrão para a nossa abordagem; taxas de detecção com limiar de 10mm (T10) para a nossa abordagem; e taxas de detecção com limiar de 15mm (T15) para a nossa abordagem e para [7].	29
Tabela 4.3	Resultados para a base BU-3DFE: erro médio, desvio padrão, taxas de detecção com limiar de 10mm (T10) e de 15mm (T15) para a nossa abordagem.	30
Tabela 4.4	Taxas de verificação sob variação de expressões faciais para a face inteira (F), a área circular em torno do nariz (C), a área elíptica em torno do nariz (E), a parte superior da face (U), e a combinação das regiões rígidas (C+E+U). . .	33
Tabela 4.5	Taxas de verificação sob variações de expressões faciais para as mesmas regiões F, C, E e U da Tabela 4.4, para a abordagem que prioriza os setores rígidos (S) e a combinação das técnicas caseadas em características faciais (C+E+U+S). . .	33

Resumo

Neste trabalho apresentamos uma abordagem para segmentação facial e detecção de características faciais em imagens de profundidade. O nosso objetivo foi o desenvolvimento de um processo automático para ser incluído em um sistema de reconhecimento facial utilizando somente a informação de profundidade como entrada. Para isto, a nossa abordagem de segmentação combina detecção de borda, agrupamento de regiões e análise de forma para extrair a região da face; e a nossa abordagem para detecção de características faciais combina classificação de curvatura de superfícies e curvas de relevo para encontrar pontos característicos no nariz e nos olhos. Os experimentos foram realizados utilizando as duas versões disponíveis da base *Face Recognition Grand Challenge* e a base BU-3DFE, buscando validar as nossas abordagens e as suas vantagens para o reconhecimento facial. Apresentamos uma análise da precisão das abordagens propostas para a segmentação e detecção de características, e comparamos os nossos resultados com outros trabalhos do estado-da-arte, publicados na literatura. Além disso, realizamos uma avaliação da influência da segmentação no reconhecimento facial e também da melhoria obtida quando as características detectadas são aplicadas para auxiliar o reconhecimento facial em imagens contendo expressões faciais.

Palavras-chave: Reconhecimento facial 3D, segmentação facial, detecção de características faciais, imagens de profundidade.

Abstract

We present a methodology for face segmentation and facial landmark detection in range images. Our goal was to develop an automatic process to be embedded in a face recognition system using only depth information as input. To this end, our segmentation approach combines edge detection, region clustering and shape analysis to extract the face region; and our landmark detection approach combines surface curvature information and relief curves to find the nose and eyes landmarks. The experiments were performed using the two available versions of the Face Recognition Grand Challenge database and the BU-3DFE database, in order to validate our proposed methodology and its advantages for 3D face recognition purposes. We present an analysis regarding the accuracy of our segmentation and landmark detection approaches. Our results were better compared to state-of-the-art works published in the literature. We also performed an evaluation regarding the influence of the segmentation process in our 3D face recognition system and analysed the improvements obtained when applying landmark-based techniques to deal with facial expressions.

Keywords: 3D face recognition, face segmentation, facial landmark detection, range images.

1 Introdução

Com o desenvolvimento das técnicas de reconhecimento facial 3D, a segmentação facial e detecção de características faciais se tornaram etapas de pré-processamento fundamentais para o desenvolvimento de sistemas de reconhecimento completamente automáticos. Essa necessidade foi observada já nos primeiros trabalhos focados no reconhecimento facial 3D, baseados em diferentes estratégias (*e.g.* registro de imagens [8, 9]; reconhecimento baseado em aparência [10–12]; e curvas de perfil [13, 14]). Em todos estes trabalhos, a face foi considerada um objeto rígido, e a mesma dificuldade foi observada: reconhecer faces contendo variações de forma. Até mesmo pequenas variações entre duas imagens de uma mesma pessoa contendo a expressão neutra causavam a diminuição das taxas de verificação. A detecção de características foi aplicada com sucesso em trabalhos mais recentes para lidar com este problema.

Chang *et al.* [15] propôs o uso de três diferentes regiões em torno do nariz durante a etapa de reconhecimento. Lu e Jain [16] propuseram a criação de modelos das deformações da face causadas por expressões faciais. Bronstein *et al.* [17] propôs uma forma de representar a face de forma invariante à expressões facial. Outros trabalhos utilizaram ideias semelhantes para aprimorar o reconhecimento facial [3, 18, 19]. Todas estas abordagens mencionadas utilizaram características faciais para atingir seus objetivos.

A classificação de curvatura de superfícies [20] tem sido aplicada com sucesso para a detecção de características faciais [15, 21, 22]. Colbry *et al.* [23] aplicou esta técnica combinada com algumas heurísticas para localizar características faciais em imagens de face contendo diferentes poses. Entretanto, esta abordagem requer uma tolerância alta a falta de precisão para obter boas taxas de localização. Recentemente, Faltemier *et al.* [24] utilizou curvas de perfil de diferentes vistas de uma imagem de entrada para encontrar a ponta do nariz em imagens não frontais de faces. Esta abordagem apresentou taxas de localização elevadas e também resultados mais precisos que a abordagem anterior. Apesar disso, esta técnica é limitada à detecção de uma única característica facial.

Para imagens frontais, Lu e Jain [5, 25] obtiveram bons resultados considerando pre-

cisão e taxas de localização através da combinação de classificação de curvatura de superfícies, curvas de relevo, modelos estatísticos da localização das características faciais e detecção de cantos em imagens de intensidade. Nesta abordagem, foram necessários o uso de imagens de intensidade e um estágio de treino para a geração do modelo estatístico. Apesar da imagem de intensidade ser facilmente adquirida, o seu uso envolve diversos problemas, entre eles variações de iluminação, pose, escala, e a possibilidade de haver um alinhamento incorreto entre as informações de profundidade e de intensidade [2].

Para realizar a detecção de características faciais, cada imagem de profundidade precisa ser previamente segmentada buscando isolar a região da face das outras partes da imagem (*i.e.* cabelo, pescoço, orelhas e roupas). Para imagens contendo apenas uma face, várias técnicas foram adotadas para a segmentação, por exemplo agrupamento de regiões [12], modelo de cor de pele [15], histograma de valores de profundidade [17], e até mesmo segmentação manual [22]. Alguns destes métodos [12, 15] utilizam a informação de intensidade para realizar a segmentação, e estão sujeitos aos mesmos problemas causados pelas variações na imagem de intensidade ou pelo desalinhamento entre as informações de profundidade e de intensidade mencionados anteriormente [2].

Outros trabalhos podem detectar mais de uma face por imagem através da utilização de técnicas de reconhecimento baseado em aparência em projeções 2D da imagem 3D [21] ou na informação de intensidade disponível [5]. Entretanto, estes métodos não extraem completamente a região da face. Em [21] somente a região em torno do nariz e dos olhos é extraída, e em [5] a segmentação exclui partes da testa e do queixo.

Neste trabalho apresentamos uma abordagem para segmentação facial e detecção de características faciais em imagens de profundidade. O nosso objetivo foi o desenvolvimento de um processo automático para ser incluído em um sistema de reconhecimento facial utilizando somente a informação de profundidade como entrada. Nossa primeira contribuição é uma abordagem de segmentação capaz de localizar e extrair faces frontais sem a necessidade de um estágio de treino. Para isto, a nossa abordagem combina detecção de borda, agrupamento de regiões e análise de forma. Esta abordagem foi mais precisa quando comparada com outras duas abordagens de segmentação [2, 5], considerando os resultados de verificação obtidos por uma técnica padrão de reconhecimento facial [26] a pelo nosso próprio sistema de reconhecimento facial 3D. Nossa segunda contribuição é uma abordagem para detecção de características faciais que combina classificação de curvatura de superfícies e curvas de relevo. Esta abordagem apresenta resultados equivalentes ou melhores quanto comparada com outros três trabalhos do estado-da-arte, presentes na literatura [5–7, 25]. Os resultados da detecção de características

foram aplicados com sucesso para aprimorar o desempenho do nosso sistema de reconhecimento facial 3D sob variações de expressão [3].

Em nossos experimentos foram utilizadas as duas versões da base *Face Recognition Grand Challenge*¹ (FRGC) e a base BU-3DFE [27], totalizando 7450 imagens de faces frontais, contendo um único indivíduo por imagem. Estas bases foram utilizadas em nossos experimentos por serem extensivamente utilizadas em pesquisas abordando análise facial 3D [2, 3, 15, 19, 28–34].

O restante deste trabalho é organizado da seguinte maneira: na Seção 2, introduzimos nossa abordagem de segmentação facial; o método de detecção de características faciais é apresentado na Seção 3; resultados experimentais são discutidos na Seção 4, seguidos pelas considerações finais na Seção 5.

¹Available at <http://www.bee-biometrics.org>

2 Segmentação facial em imagens de profundidade

Desenvolvemos uma abordagem de segmentação para extrair completamente a região da face de uma imagem de profundidade de entrada que obedeça as seguintes restrições: (1) a imagem não pode conter mais de uma face; (2) a face precisa estar entre as regiões mais próximas do dispositivo de aquisição; (3) a face deve estar posicionada frontalmente; e (4) oclusão só é permitida caso seja causada por partes do cabelo. Embora exista algumas restrições na imagem de entrada, diversas situações podem interferir no processo de segmentação, como diferentes cortes de cabelo, ou a presença de acessórios de vestuário e partes do corpo.

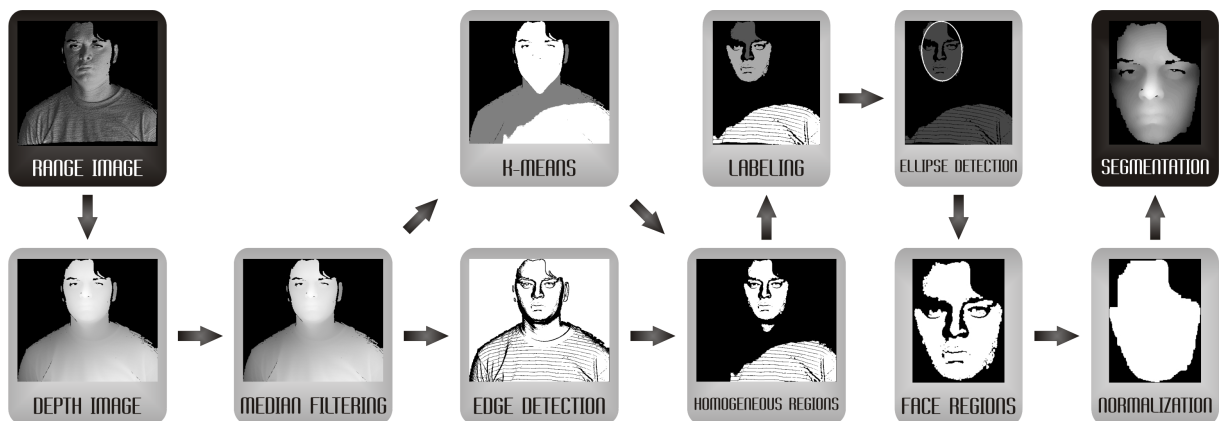


Figura 2.1: Diagrama da nossa abordagem de segmentação.

Para resolver este problema, nós desenvolvemos um algoritmo que combina detecção de bordas [35], agrupamento de regiões [36] e análise de forma. O algoritmo de segmentação possui duas etapas principais: (A) a localização de regiões homogêneas na imagem; e (B) a identificação das regiões homogêneas que pertencem à face. Um diagrama da abordagem proposta é apresentado na Fig. 2.1. Detalhes de cada estágio do processo de segmentação são fornecidos nas subseções seguintes.

2.1 Localização de regiões homogêneas

Para localizar partes homogêneas, agrupamento de regiões e detecção de bordas foram aplicados na informação de profundidade da imagem de entrada. Considerando as bases de dados utilizadas em nossos experimentos, observamos que as imagens raramente apresentam informação de profundidade no fundo da imagem, e também que a face é uma das regiões mais próximas do dispositivo de aquisição.

Primeiro, o filtro da mediana 5×5 é aplicado duas vezes na imagem de entrada para reduzir o ruído. Em seguida, o algoritmo de agrupamento K-Means [37] é aplicado na informação de profundidade para dividi-la em duas regiões (*i.e.* $K = 2$, mostradas em branco e cinza na Fig. 2.1, sendo a região em preto a área sem informação de profundidade da imagem de entrada): a região de interesse (ROI, *Region Of Interest*), composta de regiões próximas ao dispositivo de aquisição; e as demais regiões que estão muito afastadas do mesmo dispositivo e eventualmente parte do fundo da imagem. Considerando que a face sempre está incluída na região com menor profundidade (*i.e.* ROI), a área de busca pode ser reduzida, como mostrado na Fig. 2.1.

Entretanto, o agrupamento de regiões não é suficiente para isolar corretamente a face de outras regiões como cabelo, pescoço, roupas e acessórios. A detecção de bordas é então empregada para separar estas partes da região da face após a aplicação do K-Means.

Uma operação de limiarização foi empregada para a detecção de bordas em uma imagem de gradiente, que é obtida através da aplicação do operador de Sobel [38] na informação de profundidade de entrada. Um limiar global T_G foi definido baseado no ponto de inflexão do trecho decrescente do histograma de gradiente obtido utilizando todas as imagens da base FRGC v2.0 (*i.e.* o ponto de inflexão representa uma mudança significativa na curva do histograma de gradiente, e esta mudança foi considerada a mudança entre valores de gradiente de regiões homogêneas e não homogêneas). A Fig. 2.2(a) mostra o histograma de gradiente da base FRGC v2.0. O ponto de inflexão pode ser encontrado como um pico negativo na primeira derivada da curva do histograma (ver Fig. 2.2(a)), como mostrado na seguinte equação:

$$T_G = u \text{ if } h'(u) < h'(v) \forall v \neq u \quad (2.1)$$

onde $h'(x)$ é a primeira derivada da curva do histograma (*i.e.* neste trabalho, $T_G = 7$).

Imagens com alta resolução apresentam valores de gradiente menores que imagens com resoluções menores, e por este motivo o valor de T_G precisa ser normalizado de acordo

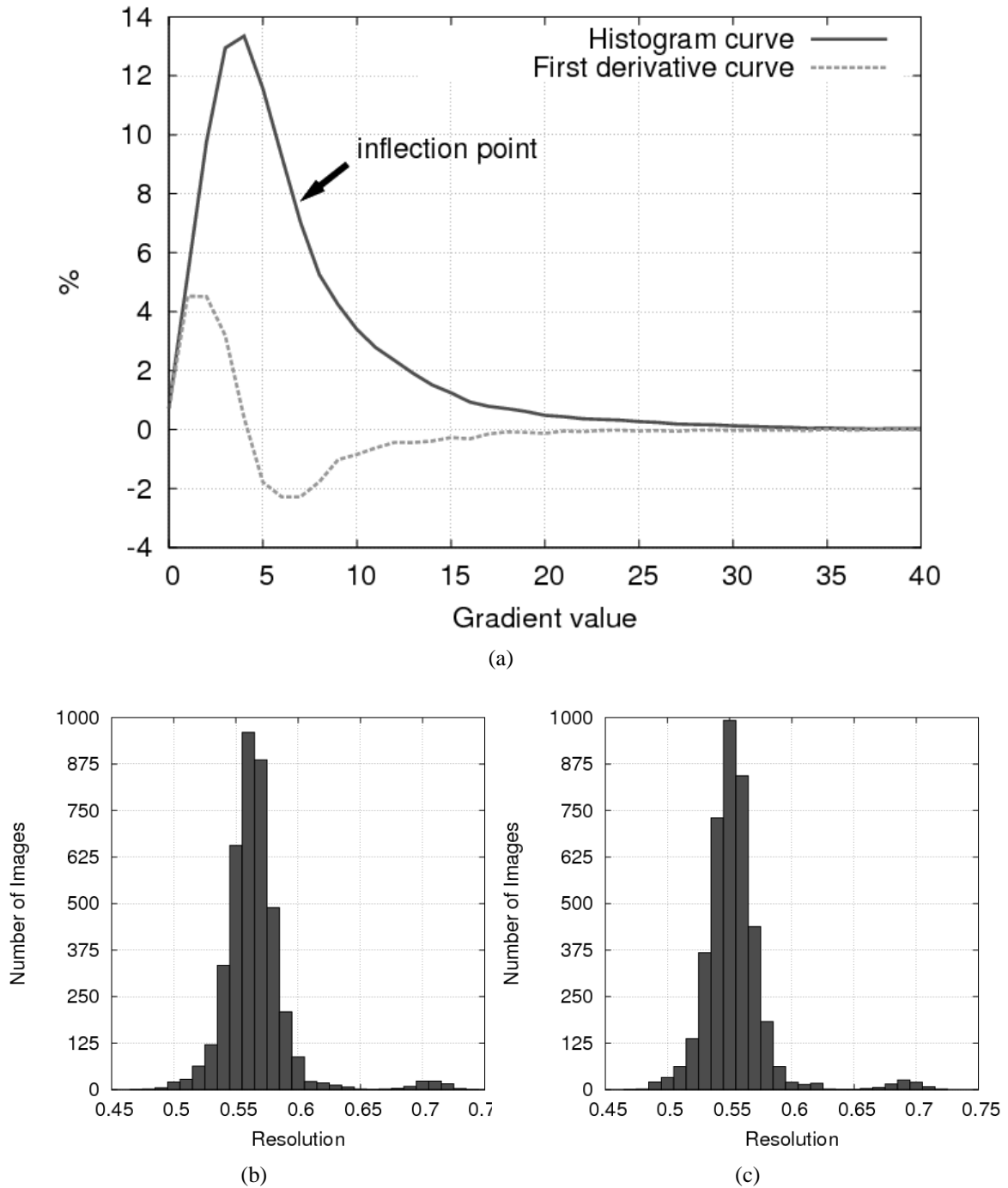


Figura 2.2: (a) Histograma de gradiente de todas as imagens da base FRGC v2.0. O ponto de inflexão representa o limiar global de gradiente; (b)-(c) histograma de valores de resolução de todas as imagens da base FRGC v2.0 utilizando o eixo y e o eixo x, respectivamente.

com a resolução da imagem. O valor de resolução r_F para uma imagem F é calculado como a distância média no eixo y entre pontos vizinhos em uma mesma coluna da imagem (*i.e.* somente o eixo y é utilizado porque ambos os eixos x e y apresentam valores similares para uma mesma imagem, como mostrado nas Figs. 2.2(b)-2.2(c)). Um valor de resolução global r_G é calculado como a média dos valores de resolução de todas as imagens da base FRGC v2.0 (ver Fig. 2.2(b),

neste trabalho, $r_G = 0.6$), e um limiar automático T_F é obtido para a imagem F através da Eq. 2.2.

$$T_F = \frac{r_F}{r_G} T_G \quad (2.2)$$

Um processo de fechamento [39] (*i.e.* duas dilatações 3×3 seguidas de uma erosão 3×3) é executado para conectar algumas linhas desconectadas resultantes do processo de limiarização e o mapa de bordas final é obtido. Após isto, o mapa de bordas é utilizado para dividir a região resultante da aplicação do K-Means combinando as duas imagens através de um E lógico, como mostrado na Fig. 2.1. A imagem resultante contém as regiões homogêneas da imagem de entrada.

2.2 Identificação de regiões da face

Neste estágio, precisamos identificar quais regiões homogêneas pertencem à face. Estas regiões são então submetidas a um processo de rotulação, e todas as regiões com tamanho inferior à 0.5% do tamanho da imagem são descartadas. Este limiar, que foi definido empiricamente, deve ser grande o suficiente para eliminar regiões indesejáveis (*e.g.* partes do cabelo), e pequeno o suficiente para nunca descartar uma região relevante. A Fig. 2.3(a) mostra um exemplo das regiões homogêneas após a rotulação.

Regiões pertencentes à face são encontradas na imagem rotulada através da busca por uma elipse, por ser considerada a forma geométrica que mais se assemelha à forma da face. Esta busca é realizada através da aplicação transformada de distância [40] nas bordas das regiões rotuladas, mostradas na Fig. 2.3, seguida da seleção da combinação de regiões que apresenta a forma mais parecida com uma elipse. Para avaliar a similaridade entre um conjunto de regiões e uma elipse nós utilizamos o Alg. 1, onde a distância entre as bordas, preenchimento da elipse e vazamento da elipse são considerados, e elipses maiores são priorizadas. A Fig. 2.4 mostra todas as combinações entre as regiões da Fig. 2.3(a). A melhor elipse, de acordo com o Alg. 1, foi encontrada na Fig. 2.4(g).

Após a detecção de elipse segue a seleção das regiões rotuladas que pertencem à face, de acordo com o Alg. 2. No passo 2 do Alg. 2, regiões próximas ao centro da elipse são selecionadas, evitando a seleção de regiões do pescoço e do cabelo, que eventualmente aparecem nas bordas da elipse. Entretanto, algumas vezes a face é dividida em pequenas regiões, e o passo 2 não é suficiente para selecionar todas estas partes. Os passos 4-7 foram desenvolvidos

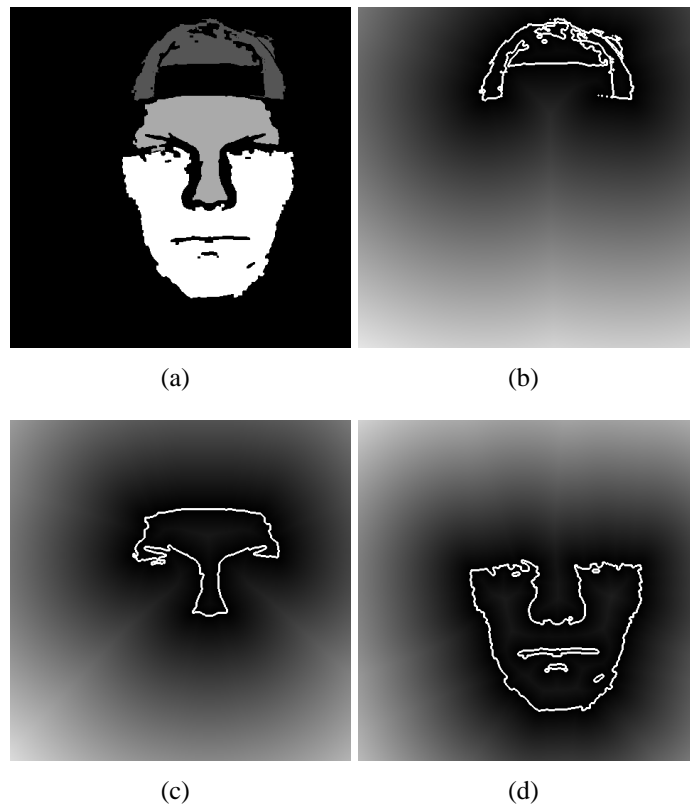


Figura 2.3: (a) Regiões homogêneas após a rotulação; (b)-(d) mapas de distância destas regiões.

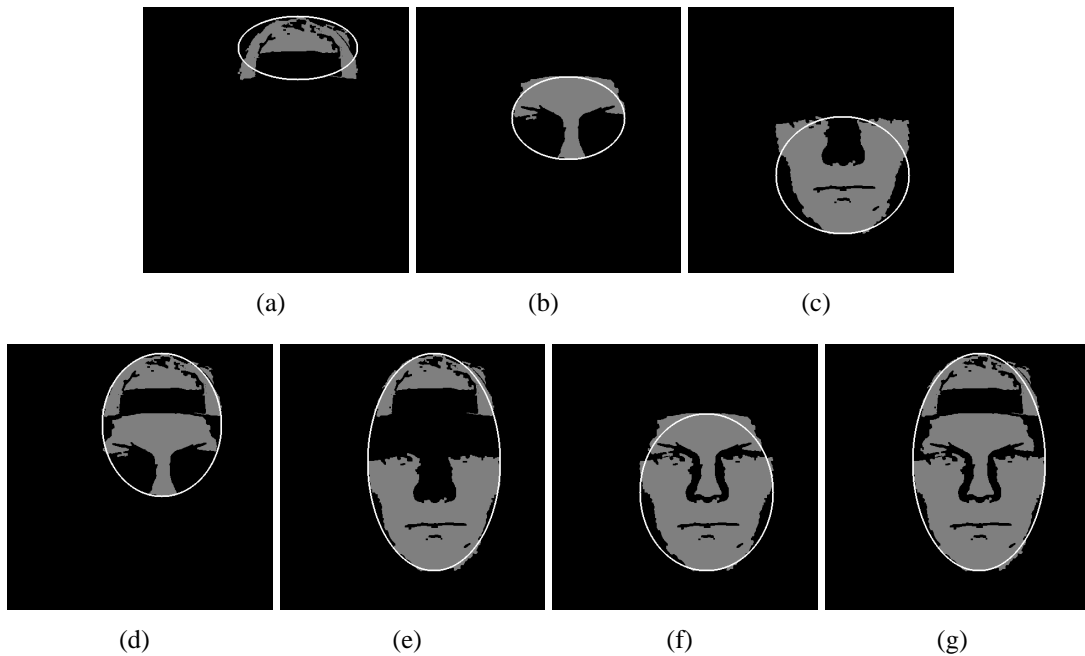


Figura 2.4: Detecção de elipse através da combinação de regiões homogêneas.

para selecionar as regiões da face neste caso. O passo 7 é aplicado para selecionar a região da testa e também evitar a seleção de partes do cabelo ou acessórios que podem estar incluídas na elipse.

Algorithm 1 Pseudo código para medir a similaridade entre regiões homogêneas e elipses.

Entrada: Um conjunto de regiões $S = \{r_1, r_2, \dots, r_N\}$

Saída: Um valor de similaridade V

- 1: Criar um mapa de distância D combinando os mapas de distância de cada região utilizando a seguinte equação para cada pixel i da imagem: $D_i = \min\{D_i^{r_1}, D_i^{r_2}, \dots, D_i^{r_N}\}$
 - 2: Obter o retângulo envolvente de S , e definir uma elipse com os valores de centro, altura e largura deste retângulo.
 - 3: Considerando P o conjunto de pixels na borda da elipse, computar a distância $d = \frac{\sum_{p \in P} D_p}{|P|}$
 - 4: Computar a fração α dos pixels não incluídos em nenhuma região que estão dentro da elipse para medir o preenchimento da elipse
 - 5: Computar a fração β dos pixels das regiões de entrada dentro do retângulo envolvente e fora da elipse para medir o vazamento da elipse
 - 6: Computar a área do retângulo envolvente A
 - 7: $V = \frac{\alpha\beta d}{A}$
-

Algorithm 2 Pseudo código para seleção de regiões da face.

Entrada: Um conjunto de regiões rotuladas $S = \{r_1, r_2, \dots, r_N\}$ e uma elipse E

Saída: Um conjunto de regiões da face F

- 1: Definir um círculo C no centro da elipse com diâmetro igual a dois terços da largura da elipse (ver Fig. 2.5(a))
 - 2: Adicionar a F todas as regiões em S com pelo menos um pixels dentro de C
 - 3: Dividir a elipse em quatro partes: a parte superior E_T ; a parte inferior E_B ; a parte esquerda E_L ; e a parte direita E_R (ver Fig. 2.5(b))
 - 4: Adicionar a F a região em S com a maior quantidade de pixels dentro de E_B
 - 5: Adicionar a F a região em S com a maior quantidade de pixels dentro de E_L
 - 6: Adicionar a F a região em S com a maior quantidade de pixels dentro de E_R
 - 7: Considerando $r_i \in S$ a região com a maior quantidade de pixels dentro de E_T , r_i será selecionada se esta também for a região com a maior quantidade de pixels na parte inferior de E_T (ver Fig. 2.5(c)), ou se tiver mais que o dobro da quantidade de pixels da região mais comum na parte inferior de E_T
-

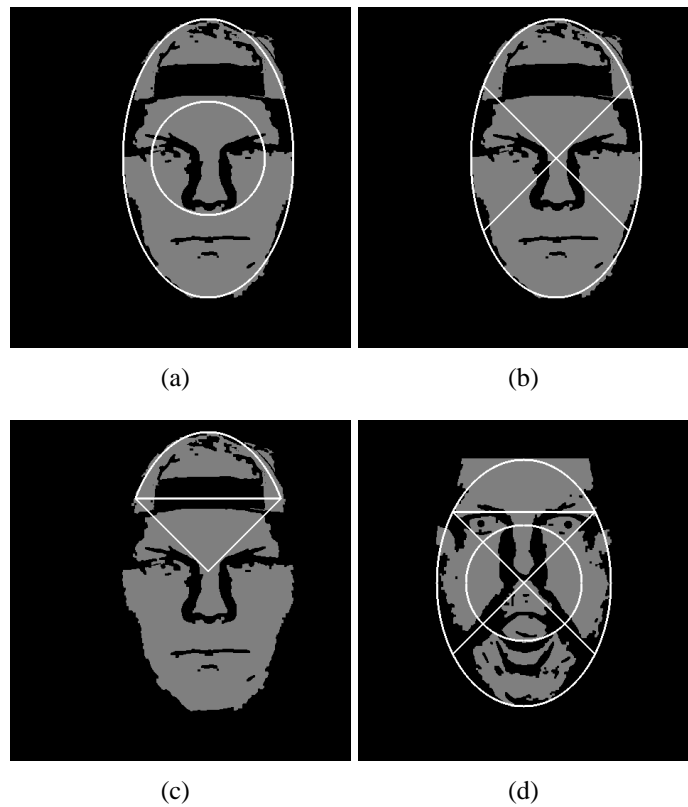


Figura 2.5: (a)-(c) Regiões incluídas na elipse detectada utilizadas para localizar a face; e (d) um exemplo de face dividida por causa do bigode.

As regiões selecionadas como face são representadas por uma imagem binária que indica a localização da face na imagem de entrada. Esta imagem pode conter espaços entre as regiões, que são preenchidos através da ligação de pontos desconectados em uma mesma coluna ou em uma mesma linha da imagem. Após isto, realizamos um E lógico entre a imagem binária resultante e a imagem de profundidade de entrada para obter a segmentação final da região da face.

3 Detecção de características faciais

A detecção de características faciais é uma etapa importante para otimizar algumas tarefas relacionadas ao reconhecimento; por exemplo registro de faces, estimativa de pose, e localização de regiões rígidas da face [15].

Estamos interessados em um pequeno conjunto de características faciais: ponta do nariz e cantos do nariz e dos olhos. Estas características foram escolhidas porque elas permitem uma extração precisa das regiões rígidas da face, como o nariz e a testa. Para localizá-las, nós combinamos curvas de relevo [24, 41] obtidas da informação de profundidade e também da imagem resultante da classificação de curvatura de superfícies [4, 15]. Curvas de relevo [24, 41] são curvas extraídas de imagens para representar alguma propriedade específica da imagem (*e.g.* o valor máximo de profundidade cada linha da imagem da face representa a linha de perfil da face). Um diagrama da abordagem de detecção de características é apresentado na Fig. 3.1.

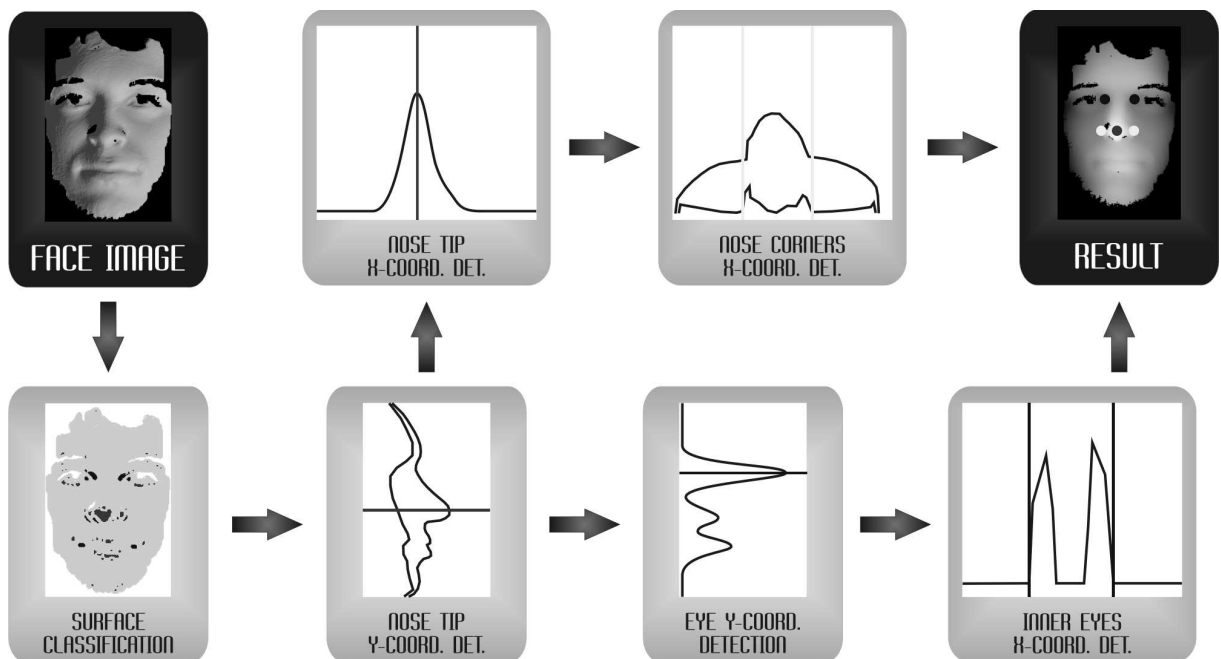


Figura 3.1: Diagrama da nossa abordagem de detecção de características faciais.

3.1 Classificação de curvatura de superfícies

Como mostrado por Chang *et al.* [42], através da análise da informação de curvatura de superfícies podemos observar que algumas regiões da face (*e.g.* ponta do nariz e cantos internos dos olhos) apresentam o mesmo tipo de curvatura em imagens de diferentes pessoas, e não são afetadas por expressões faciais, como mostrado na Fig. 3.2.

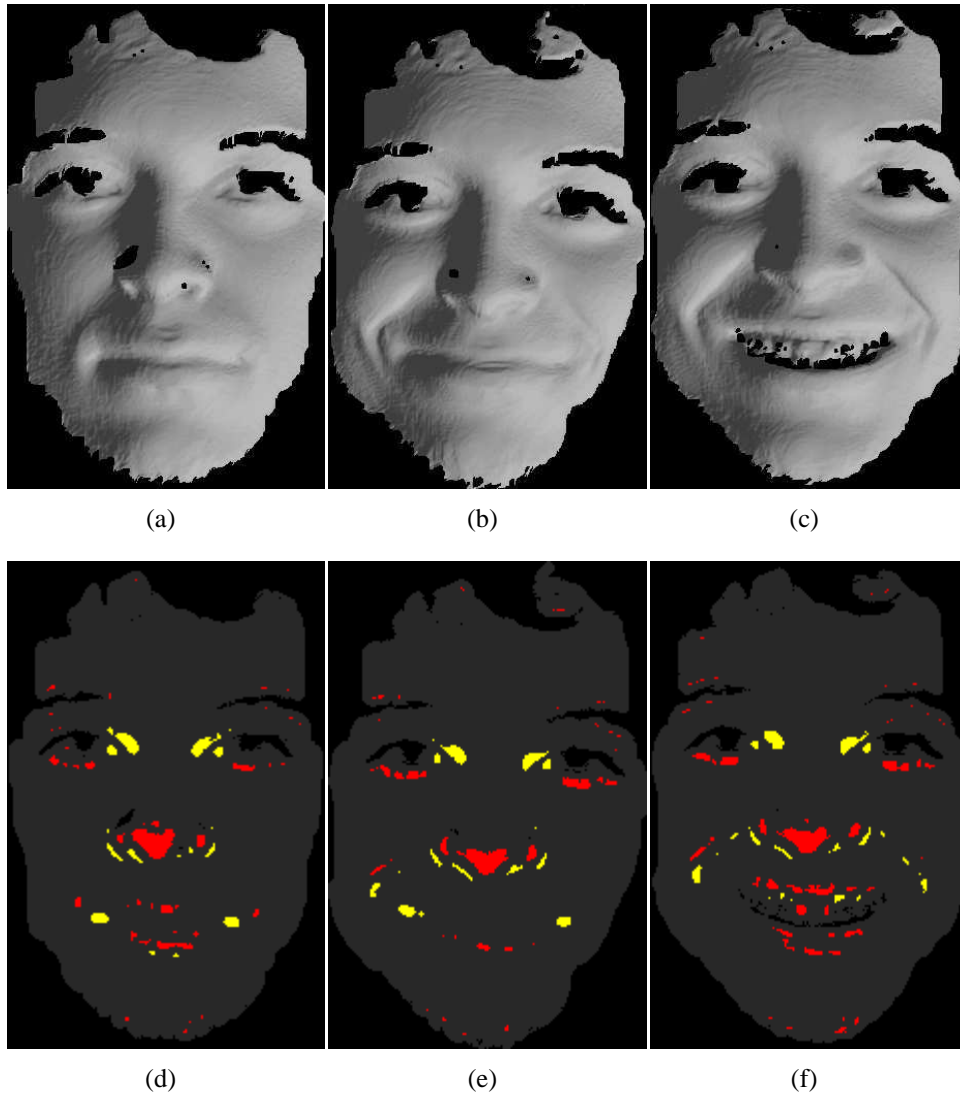


Figura 3.2: Exemplo da classificação de superfícies aplicada em imagens de uma mesma pessoa apresentando diferentes expressões: (a), (b) e (c) são os resultados da segmentação; (d), (e) e (f) mostram a classificação de superfícies das imagens (a), (b) e (c), respectivamente. Regiões em vermelho mostram picos na superfície e regiões em amarelo mostram poços.

Para calcular o tipo de curvatura para cada ponto de uma imagem de face, uma superfície local é definida através de uma vizinhança $N \times N$, assim como em [43] (*i.e.* $N = 11$). Então nós aplicamos uma técnica de mínimos quadrados para calcular os coeficientes de uma aproximação biquadrática da superfície, e estimamos os valores de curvatura Gaussiana (K)

e de curvatura média (H) utilizando as derivadas parciais da aproximação [4]. Uma vez que os valores K e H tenham sido obtidos, podemos classificar a superfície local de acordo com a Tabela 3.1 [4].

Tabela 3.1: Classificação dos tipos de superfícies baseada nos valores de curvatura Gaussiana e curvatura média [4].

	$H < 0$	$H = 0$	$H > 0$
$K < 0$	pico	-	poço
$K = 0$	cume	plano	vale
$K > 0$	sela em cume	mínimo	sela em vale

Depois de classificar todos os pontos de uma imagem de entrada, dois tipos de superfícies são utilizados em nossa abordagem de localização de características faciais: picos e poços. Os cantos dos olhos apresentam superfícies em forma de poço, e a ponta do nariz apresenta superfícies em forma de pico, como mostrado na Fig. 3.2. Os cantos do nariz podem apresentar diferentes classificações sob diferentes expressões faciais, então outras informações extraídas da face foram empregada para encontrá-los.

3.2 Detecção de características

A detecção de características possui dois estágios: primeiro nós encontramos a coordenada y da ponta do nariz e dos olhos, a depois encontramos a coordenada x da ponta e cantos do nariz e dos cantos internos dos olhos.

Para encontrar a coordenada y do nariz, duas projeções da informação de profundidade no eixo y [41] são geradas, chamadas curva de perfil e curva mediana. Para gerar estas curvas, determinamos o valor de profundidade máximo (*i.e.* curva de perfil) e o valor mediano (*i.e.* curva mediana) de cada conjunto de pontos na imagem de entrada com a mesma coordenada y . Calculamos então a diferença entre estas curvas e procuramos pela coordenada y com o maior valor de diferença. Então, assumimos que o pico local na curva de perfil mais próximo desta coordenada y corresponde à localização do nariz. A Fig. 3.3 mostra um exemplo de curva de perfil e seus picos locais, de curva mediana e da diferença entre as curva de perfil e mediana. A coordenada y com o valor máximo de diferença e pico local escolhido como ponta do nariz também são mostrados na Fig. 3.3.

Em seguida, outra projeção no eixo y é gerada utilizando a imagem de curvatura para localizar a coordenada y dos olhos. Para isto, calculamos a porcentagem de pontos classificados como pico em cada conjunto de pontos com a mesma coordenada y . Para uma face, esta

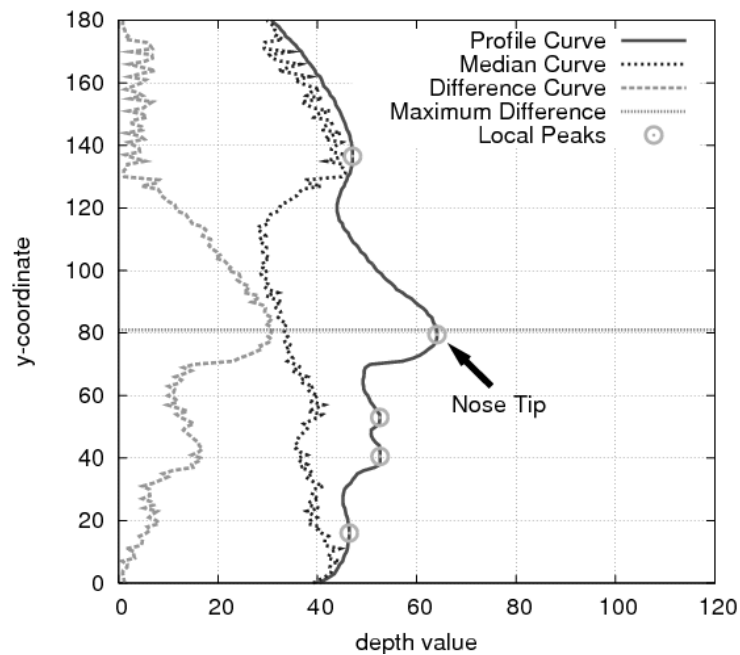


Figura 3.3: Exemplo de detecção da coordenada y da ponta do nariz.

projeção apresenta três picos, representando os olhos, a base do nariz e a boca. Como a coordenada y da ponta do nariz já é conhecida, podemos considerar o pico mais próximo à ponta do nariz como sendo a coordenada y base do nariz, o pico superior como sendo a coordenada y dos olhos, e o pico inferior a coordenada y da boca (ver Fig. 3.4).

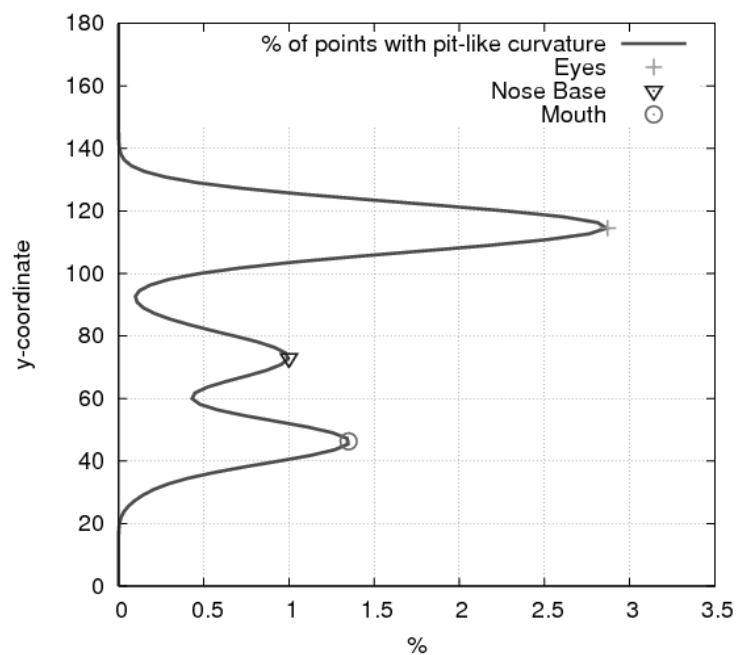


Figura 3.4: Exemplo da detecção da coordenada y dos olhos, base do nariz e boca.

Uma vez que as coordenadas y da ponta do nariz e dos olhos sejam conhecidas, com-

putamos projeções no eixo x para localizar as coordenadas x da ponta e cantos do nariz, e dos cantos internos dos olhos. Para a ponta do nariz, a projeção no eixo x é obtida através do cálculo da porcentagem de pontos em uma mesma coluna considerando um conjunto de linhas vizinhas à coordenada y da ponta do nariz. A coordenada x da ponta do nariz pode ser encontrada como um pico nesta projeção, como mostrado na Fig. 3.5.

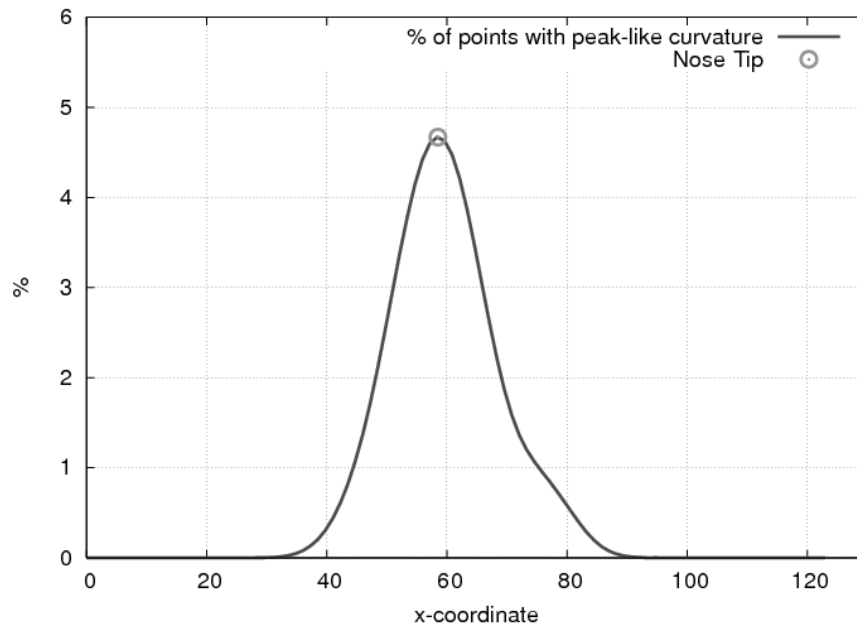


Figura 3.5: Exemplo da detecção da coordenada x da ponta do nariz.

A partir da ponta do nariz, podemos encontrar os cantos do nariz como sendo os pontos de máxima variação na curva de perfil horizontal. Esta curva é uma projeção no eixo x que representa um conjunto de pontos com a mesma coordenada y, neste caso, a coordenada y da ponta do nariz. Para detectar os cantos do nariz, nós calculamos a informação de gradiente desta curva buscando identificar um pico em cada lado da ponta do nariz, como mostrado na Fig. 3.6.

Para encontrar as coordenadas x dos cantos dos olhos, nós geramos uma projeção da imagem de curvatura no eixo x através do cálculo da porcentagem de pontos classificados como poço para cada coluna de um conjunto de linhas vizinhas centralizadas na coordenada y dos olhos. A coordenada x do canto do olho esquerdo está localizada no início do primeiro pico nesta projeção, e a coordenada x do canto do olho direito está localizada no final do segundo pico nesta mesma projeção, como mostrado na Fig. 3.7.

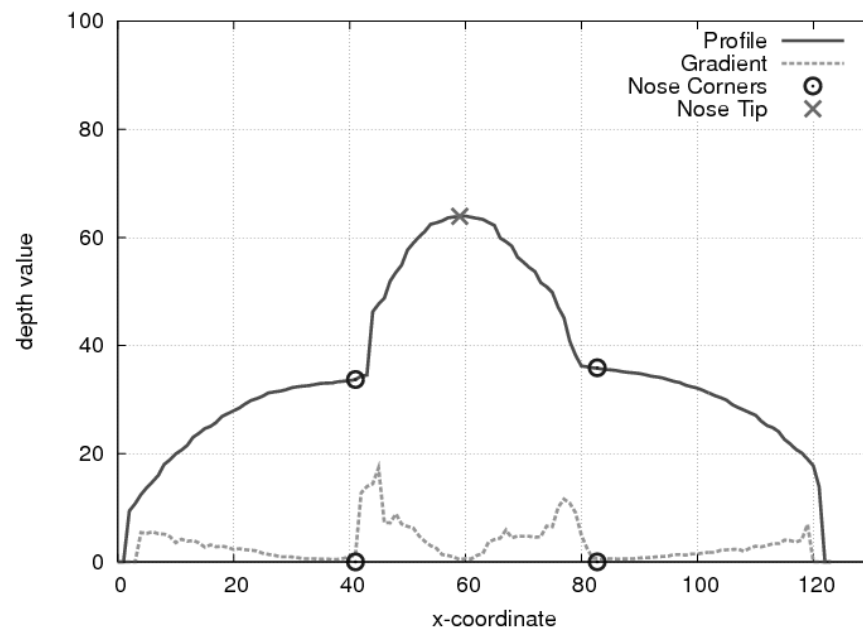


Figura 3.6: Exemplo da detecção dos cantos do nariz.

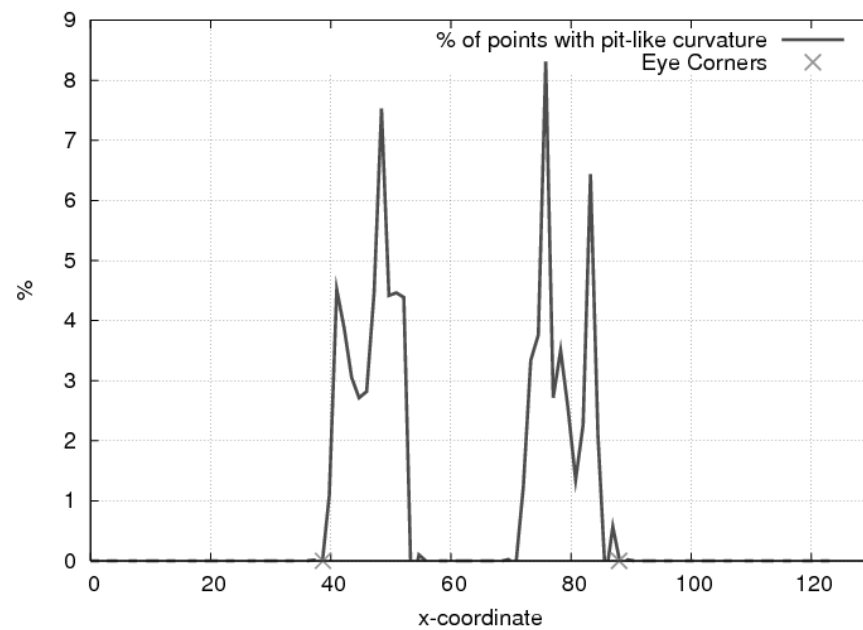


Figura 3.7: Exemplo da detecção dos cantos dos olhos.

4 Resultados experimentais

Os experimentos apresentados nesta seção foram projetados para avaliar a precisão das nossas abordagens para segmentação facial e detecção de características, a também para analisar a influência das mesmas em um sistema de reconhecimento facial 3D. Três bases de dados foram utilizadas nestes experimentos. A base FRGC versões 1.0 e 2.0, com um total de 4950 imagens de 557 indivíduos, contendo variações de expressões faciais, resolução, pose e outras características, como diferentes cortes de cabelo. As imagens possuem 640×480 pixels e foram adquiridas pelo *scanner laser* Minolta Vivid 910; a e base BU-3DFE [27], com um total de 2500 imagens de 100 indivíduos apresentando principalmente variações de expressões faciais. Apenas malhas de triângulos são disponibilizadas nos dados originais desta base, então nós convertemos os modelos para imagens de profundidade frontalmente posicionadas utilizando amostragem de pontos considerando uma resolução de $0.6mm$.

Todos os experimentos de reconhecimento foram realizados utilizando-se um método padrão de comparação baseado no algoritmo *Iterative Closest Point* (ICP) e também o nosso sistema de reconhecimento facial 3D [3] baseado nas técnicas *Simulated Annealing* (SA) e *Surface Interpenetration Measure* (SIM). O ICP foi empregado em nossos experimentos porque muitos trabalhos voltados para o reconhecimento na literatura são baseados nele. Em nossa implementação, uma solução inicial é obtida através do alinhamento dos centros de massa das duas imagens de face, então um alinhamento grosseiro é realizado com o ICP e 5% dos pontos da imagem a ser reconhecida. A seguir, um alinhamento mais preciso é obtido com o aumento da porcentagem de pontos utilizada para 25%. O valor de similaridade final é obtido através da aplicação da medida robusta *M-estimator Sample Consensus* (MSAC) sobre o erro entre pontos correspondentes na área de sobreposição das imagens alinhadas. O limiar para o MSAC foi definido em $3.0mm$. Da mesma maneira, em nosso sistema de reconhecimento uma solução inicial é obtida através do alinhamento dos centros de massa das duas imagens de face. Um alinhamento grosseiro é realizado utilizado uma busca com o SA para minimizar o MSAC, e então um alinhamento mais preciso é obtido utilizando uma busca com o SA para maximizar o valor da SIM [3].

4.1 Resultados da segmentação

A nossa abordagem de segmentação extraiu corretamente a região da face em 99.3% das imagens da base FRGC v1.0 (936 de 943), em 99.7% das imagens da base FRGC v2.0 (3993 de 4007) e em 96.1% das imagens da base BU-3DFE (2403 de 2500), como confirmado por uma inspeção visual dos resultados (*i.e.* faces contendo regiões irrelevantes ou que não foram extraídas por completo foram consideradas segmentações incorretas). Os erros foram causados principalmente por regiões do pescoço, cabelo ou acessórios que não foram separados da face pela etapa de detecção de bordas. Alguns exemplos são mostrados na Fig. 4.1. Nas Figs. 4.1(a)-4.1(e), regiões irrelevantes não foram separadas da face, e na Fig. 4.1(f), a região da testa não foi selecionada como face. O tempo médio para segmentar uma face é de um segundo em um Pentium D 3.4 GHz. A Fig. 4.2 mostra alguns exemplos da segmentação facial utilizando a abordagem proposta.

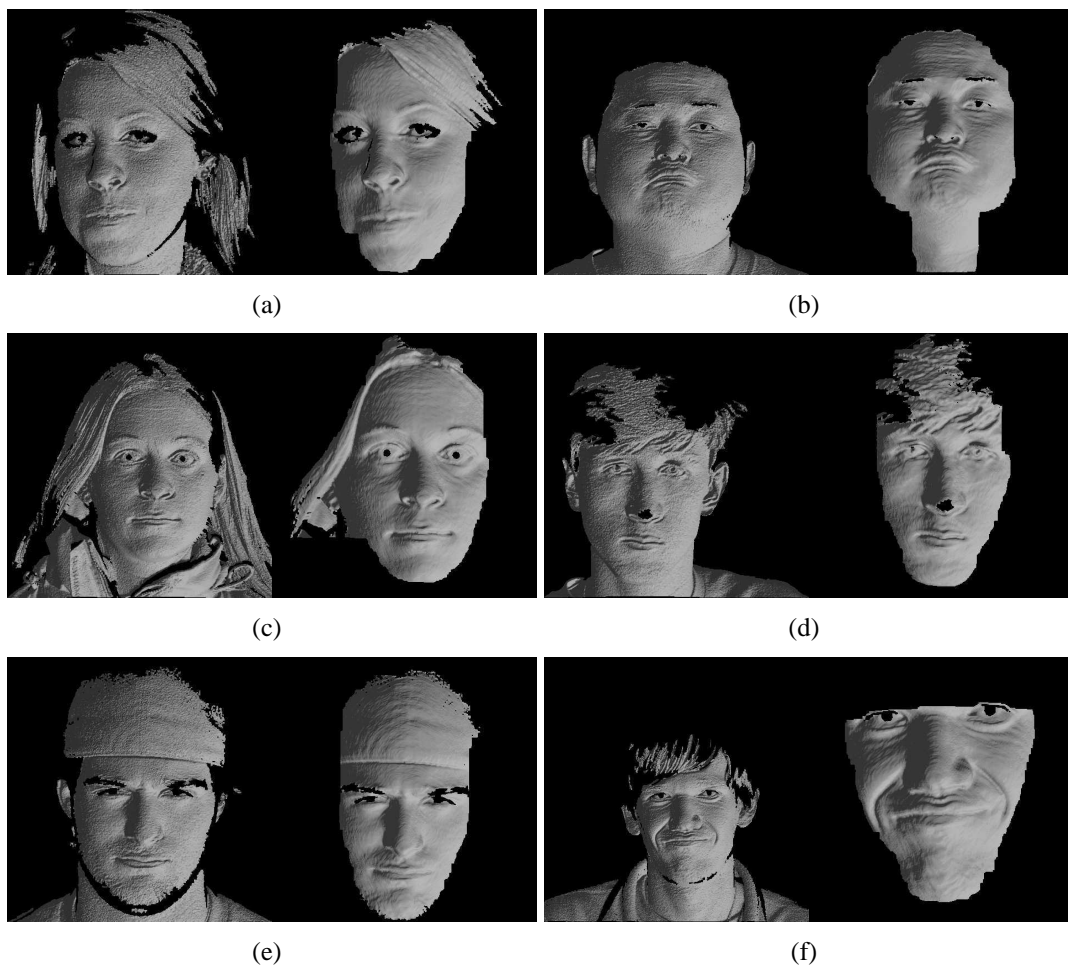


Figura 4.1: Segmentações incorretas causadas por (a)-(e) regiões irrelevantes selecionadas como face e (f) partes da face não selecionadas.

Para avaliar a influência dos resultados da segmentação em sistemas de reconheci-

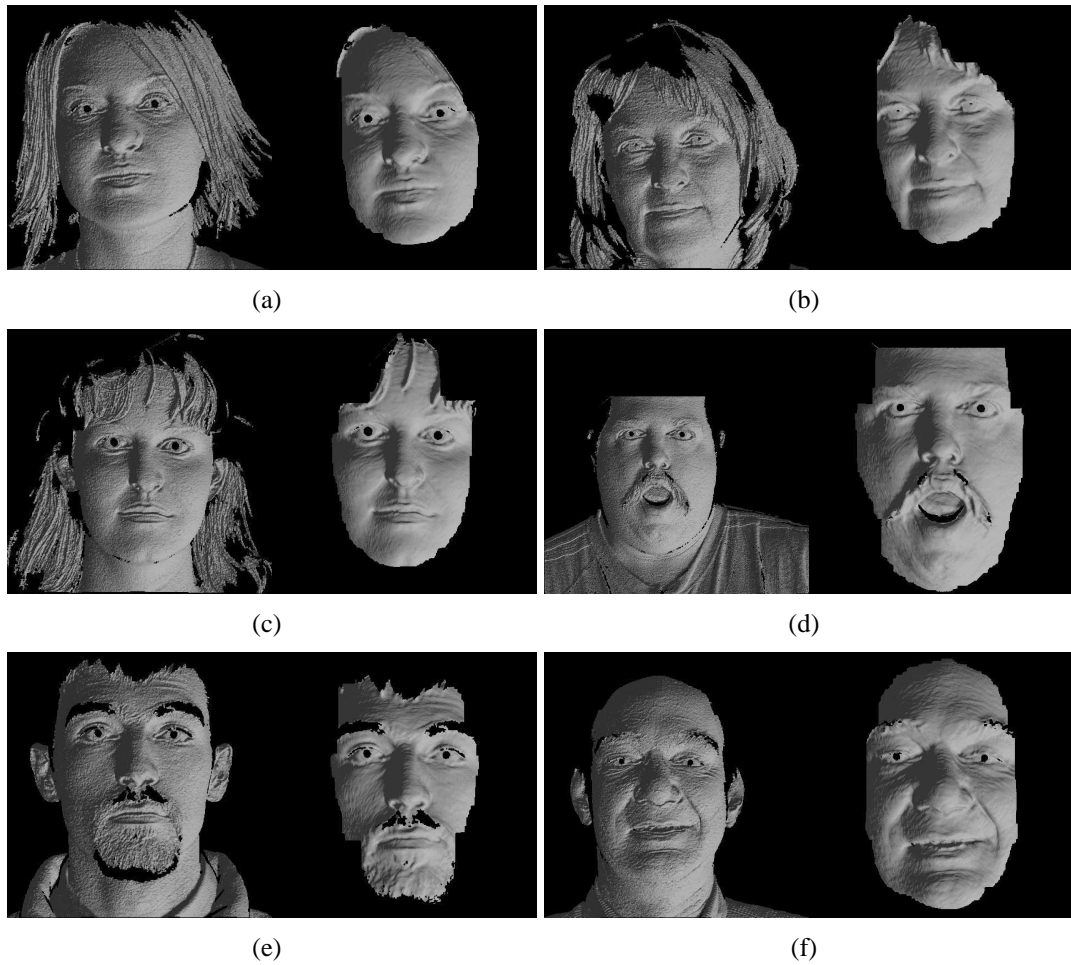


Figura 4.2: Exemplos da segmentação facial utilizando a abordagem proposta em imagens com diferentes dificuldades: (a)-(c) cabelo; (d)-(e) barba e bigode; e (f) expressão facial.

mento facial 3D, nós comparamos nossa abordagem com outras duas técnicas de segmentação [2, 5]. A primeira é uma técnica de segmentação baseada na detecção da face através de um classificador em cascata desenvolvido por Viola e Jones [1], disponível na biblioteca *OpenCV*¹, que usa a imagem de intensidade para localizar a face. O fundo da imagem de intensidade foi removido utilizando-se o mapa de profundidade, como proposto em [5], para reduzir a área de busca e o número de detecções falsas. A região da face corresponde ao conjunto de pontos localizados dentro do quadrado detectado, como mostrado na Fig. 4.3. Alguns exemplos do resultado desta técnica são mostrados na Fig. 4.4. A segunda abordagem foi proposta por Mian *et al.* [2], e consiste em extrair uma região esférica com um raio de $80mm$ em torno da ponta do nariz, como mostrado na Fig. 4.5(a). Exemplos da segmentação final obtida por esta abordagem são mostradas nas Figs. 4.5(b)-4.5(d).

Para esta comparação foram utilizadas 933 imagens da base FRGC v2.0 sem ruído e com a expressão neutra, eliminando assim problemas causados por variações de expressão

¹Available at <http://www.intel.com/technology/computing/opencv>

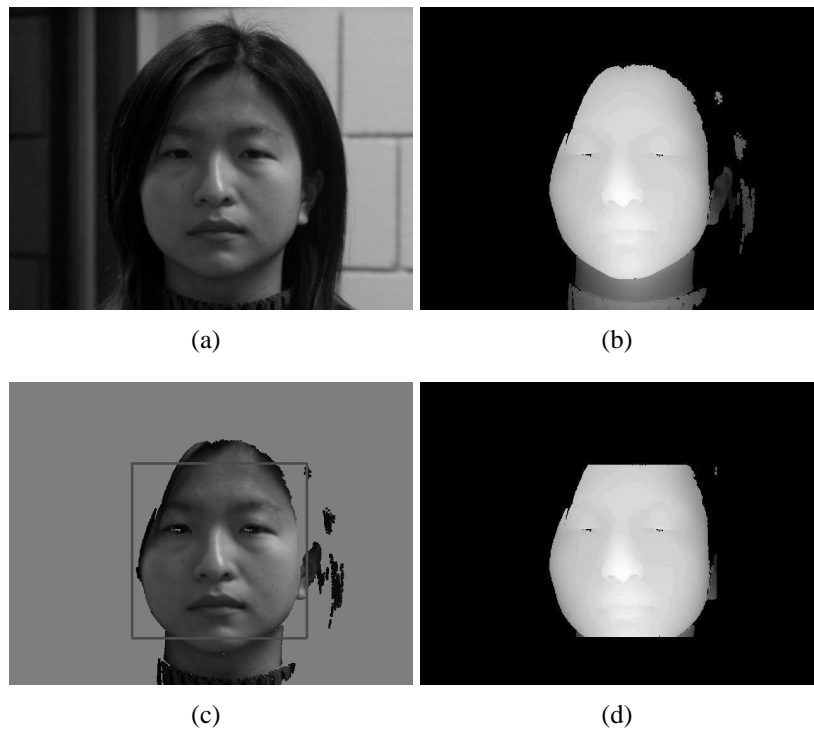


Figura 4.3: Exemplo da segmentação facial através de uma técnica padrão [1]: (a) imagem de intensidade, (b) mapa de profundidade, (c) imagem de intensidade após remoção do fundo, e (d) resultado da extração facial.



Figura 4.4: Exemplos da segmentação facial obtidas pela técnica padrão de segmentação.

e de ruído. Com esta configuração, a face inteira é relevante ao reconhecimento facial, e a segmentação que maximizar a área de face extraída e minimizar a quantidade de regiões irrele-

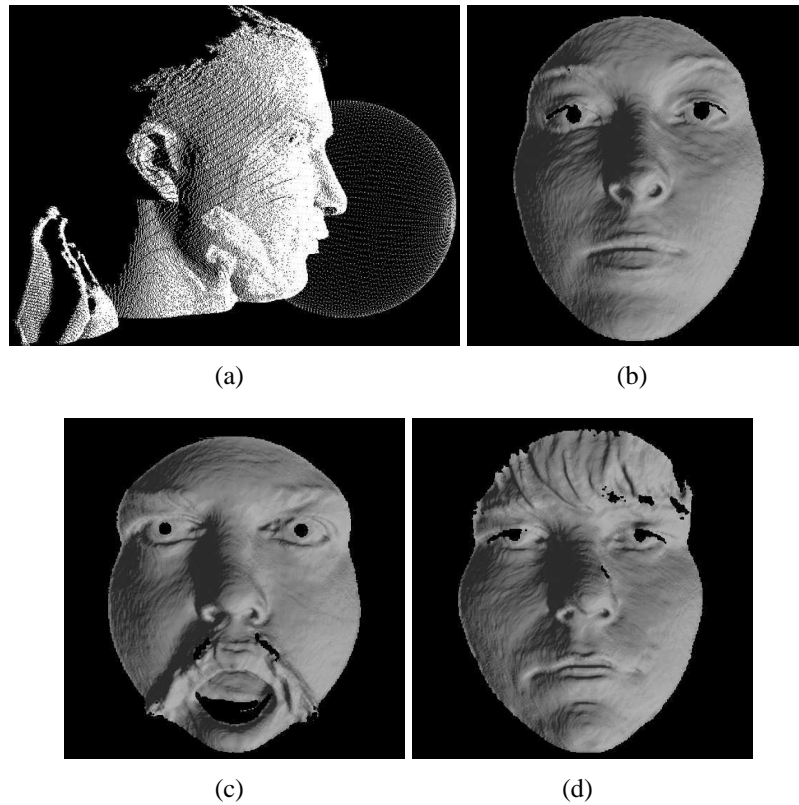


Figura 4.5: (a) Ilustração da abordagem proposta por Mian (imagem retirada de [2]), e (b)-(d) resultados obtidos por esta abordagem.

vantes apresentará as melhores taxas de verificação. Todas estas imagens neutras e sem ruído foram então segmentadas pelas três técnicas consideradas. Detecções falsas na técnica padrão foram eliminadas manualmente, e as posições da ponta do nariz utilizadas na abordagem de Mian foram marcadas manualmente.

Em seguida, nós calculamos o valor de similaridade para cada combinação de duas imagens segmentadas por uma mesma abordagem (864578 combinações entre imagens de indivíduos diferentes e 4978 entre imagens do mesmo indivíduo), e a Fig. 4.6 mostra as taxas de verificação para as três abordagens. Como pode ser observado, a nossa abordagem apresentou melhores taxas de verificação em comparação às outras abordagens.

4.2 Resultados da detecção de característica

Após a segmentação, as faces são submetidas à nossa abordagem de detecção de características sem nenhum ajuste (*i.e.* segmentações incorretas não são retificadas). As características faciais foram manualmente marcadas por duas pessoas diferentes para avaliar o erro de localização, e as posições de referência de cada característica foi definida como a média das

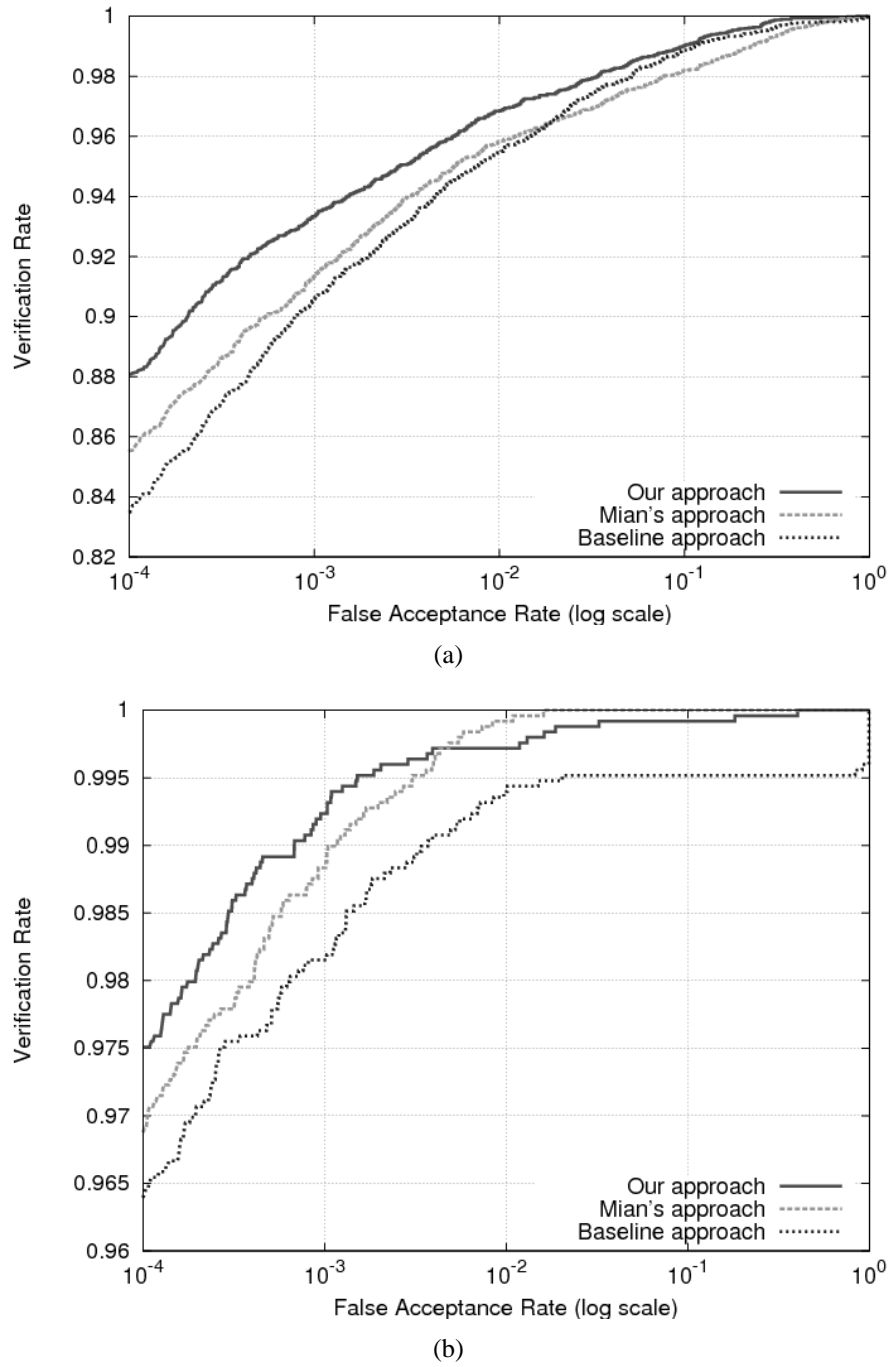


Figura 4.6: Comparação entre abordagens de segmentação utilizando (a) ICP+MSAC e (b) SA+SIM.

marcações manuais.

O erro de localização foi estimado através da distância Euclidiana entre a posição de referência e a posição obtida pela nossa abordagem de detecção. Para mostrar a melhora em termos de precisão, algumas estatísticas do erro de localização são apresentadas nas Tabelas 4.1-4.3, junto com resultados obtidos por Lu e Jain [5, 25], Romero-Huertas e Pears [7] e Yu e Moon [6]. A base FRGC v2.0 foi utilizada nos experimentos apresentados em [7] e a base

FRGC v1.0 foi utilizada em [5, 6]. O erro médio e o desvio padrão foram obtidos considerando o erro de localização entre as posições de referência e os resultados da nossa abordagem automática de detecção. As taxas de detecção foram obtidas através do cálculo da porcentagem de características detectadas com o erro de localização menor que um limiar estipulado. Como pode ser observado, a nossa abordagem apresenta valores baixos de erro médio e desvio padrão, e consequentemente a ponta do nariz e os cantos dos olhos foram detectados corretamente em mais imagens que outras abordagens quando o limiar de tolerância é baixo.

Tabela 4.1: Resultados para a base FRGC v1.0: erro médio e desvio padrão para a nossa abordagem e também para [5, 6]; taxas de detecção com limiar de 10mm (T10) para a nossa abordagem e para [6]; e taxas de detecção com limiar de 15mm (T15) para a nossa abordagem. PN=ponta do nariz; OE=olho esquerdo; OD=olho direito; NE=canto esquerdo do nariz; ND=canto direito do nariz.

Características	PN	OE	OD	NE	ND
Erro médio (mm)	2.69	4.26	4.15	6.16	7.22
[5]-Erro médio (mm)	5.00	5.70	6.00	N.A.	N.A.
[6]-Erro médio (mm)	2.18	5.59	4.74	N.A.	N.A.
Desvio padrão (mm)	2.14	3.51	3.13	3.09	2.76
[5]-Desvio padrão (mm)	2.40	3.00	3.30	N.A.	N.A.
[6]-Desvio padrão (mm)	6.83	16.08	9.76	N.A.	N.A.
T10	99.89	97.14	98.52	90.56	90.03
[6]-T10	99.50	98.50	98.50	N.A.	N.A.
T15	99.89	99.47	99.68	99.79	99.47

Tabela 4.2: Resultados para a base FRGC v2.0: erro médio e desvio padrão para a nossa abordagem; taxas de detecção com limiar de 10mm (T10) para a nossa abordagem; e taxas de detecção com limiar de 15mm (T15) para a nossa abordagem e para [7].

Características	PN	OE	OD	NE	ND
Erro médio (mm)	2.73	3.69	3.35	4.83	5.84
Desvio padrão (mm)	1.39	2.26	2.33	2.03	1.73
T10	99.95	98.88	99.15	99.55	99.15
T15	99.95	99.65	99.63	99.95	99.95
[7]-T15	90.00	90.00	90.00	N.A.	N.A.

Computamos o histograma de erro de localização (LDH, *Localization Displacement Histogram*) para cada característica utilizando as imagens da base FRGC v2.0. Os gráficos de LDH para ponta do nariz, cantos internos dos olhos e cantos do nariz são mostrados na Fig. 4.7. O tempo médio para localizar as características faciais em uma face segmentada é de

Tabela 4.3: Resultados para a base BU-3DFE: erro médio, desvio padrão, taxas de detecção com limiar de 10mm (T10) e de 15mm (T15) para a nossa abordagem.

Características	PN	OE	OD	NE	ND
Erro médio (mm)	1.87	6.33	6.33	6.66	6.49
Desvio padrão (mm)	1.12	4.82	5.04	3.36	3.40
T10	99.88	84.32	82.40	83.56	83.56
T15	99.88	94.60	93.20	99.08	99.52

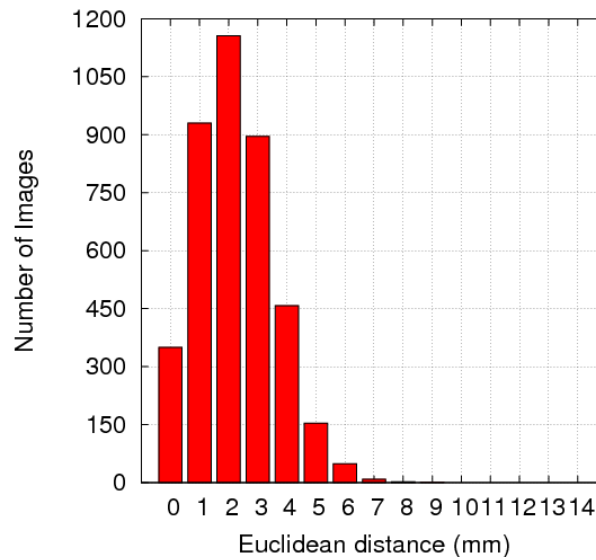
0.3 segundos em um Pentium D 3.4 GHz. Alguns resultados da detecção de características são mostrados na Fig. 4.8.

Após detectadas, estas características foram utilizadas para tornar sistemas de reconhecimento facial 3D capazes de reconhecer imagens sob variações causadas por expressões faciais. Para isto, realizamos a comparação entre todas as imagens neutras da base FRGC v2.0, assim como todas as comparações entre imagens neutras e não-neutras. Este experimento reproduz um sistema de reconhecimento facial genuíno, onde faces não controladas são comparadas com faces controladas (*i.e.* neutras e sem ruído) previamente adquiridas e armazenadas em uma base.

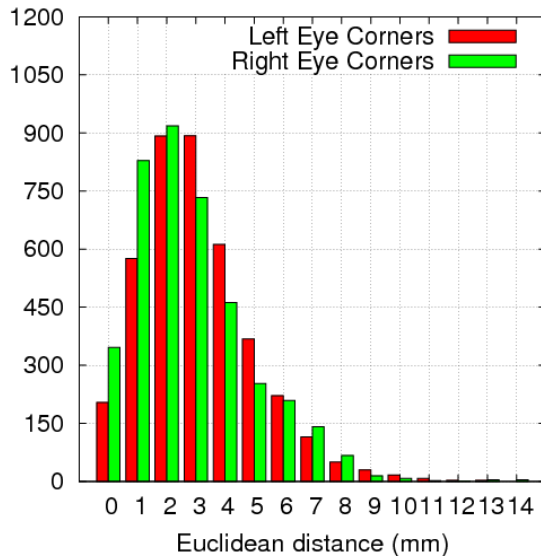
Para aprimorar o reconhecimento, duas técnicas diferentes baseadas em características faciais foram consideradas: a primeira extrai diferentes regiões da face que são menos afetadas por variações causadas por expressões faciais (*i.e.* o nariz e a testa, mostrados nas Figs. 4.9(a), 4.9(b) and 4.9(c)); e a segunda utiliza a face inteira, mas a face é dividida em nove setores e os setores contendo regiões invariantes, destacadas na Fig. 4.9(d), são priorizadas durante o processo de reconhecimento [3].

As faces foram comparadas utilizando diferentes regiões: a face inteira, a região do nariz, a parte superior da face, e a face inteira com setores priorizados. Os resultados de reconhecimento para o ICP e para o nosso sistema de reconhecimento são mostrados separadamente nas Tabelas 4.4 e 4.5.

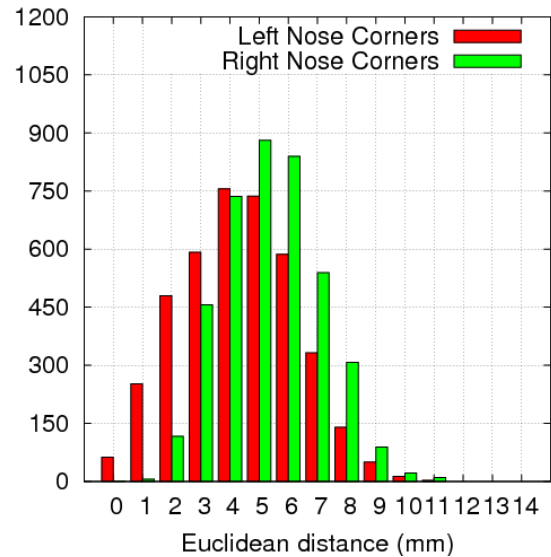
A Tabela 4.4 mostra as taxas de verificação para o sistema de reconhecimento baseado em ICP, considerando uma taxa de falsa aceitação (FAR, *False Acceptance Rate*) de 10^{-3} , para a face inteira, a área circular em torno do nariz (ver Fig. 4.9(a)) [15], a área elíptica em torno do nariz (ver Fig. 4.9(b)) [15], a parte superior da face (Fig. 4.9(c)) [44], e a combinação das regiões rígida pela regra da soma [45]. Para cada uma, diferentes taxas foram obtidas utilizando faces com diferentes expressões: neutra (NEU), feliz (HAP), boca aberta (OPM), franzida (FRO), nojo (DIS), surpresa (SUR), triste (SAD), bochechas infladas (PUF) e outras (OTH). Algumas destas expressões possuem nível baixo ou alto de intensidade, denominados 1 e 2, respectiva-



(a)



(b)



(c)

Figura 4.7: Avaliação da detecção de características: LDH para (a) ponta do nariz, (b) cantos internos dos olhos e (c) cantos do nariz.

mente. As taxas de verificação obtidas quando a face inteira é utilizada são afetadas pelas expressões faciais, e as técnicas baseadas em características faciais apresentaram diferentes comportamentos de acordo com as variações apresentadas na face. As regiões circulares e elípticas em torno do nariz são afetadas por expressões que causam variações próximas à região do nariz (*e.g.* nojo e bochechas infladas), e a parte superior da face apresenta problemas com expressões que causam variações nas regiões do nariz e da testa (*e.g.* franzida, nojo e triste). A combinação das regiões rígidas apresentou melhores resultados do que a utilização de uma única região em mais da metade das expressões analisadas, e também uma maior estabilidade para diferentes expressões.

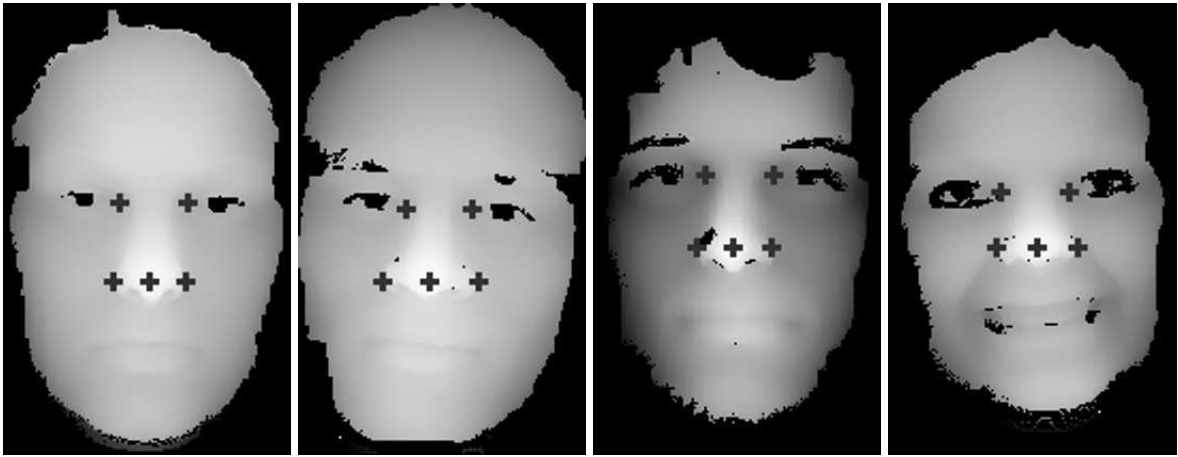


Figura 4.8: Resultados da nossa abordagem de detecção de características faciais.

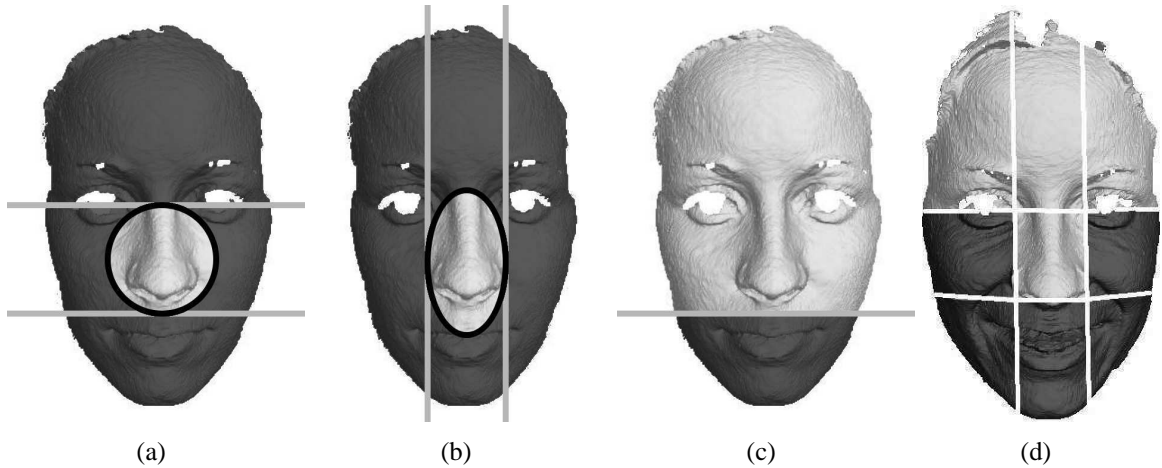


Figura 4.9: Exemplo das técnicas para aprimorar o reconhecimento baseadas em características (imagens retiradas de [3]): (a) área circular em torno do nariz, (b) área elíptica em torno do nariz, (c) parte superior da face, e (d) abordagem com setores priorizados.

A Tabela 4.5 mostra as taxas de verificação para nosso sistema de reconhecimento facial, considerando FAR de 10^{-3} , para a face inteira, a área circular em torno do nariz, a área elíptica em torno do nariz, a parte superior da face, a abordagem que prioriza os setores rígidos (Fig. 4.9(d)) [3] e a combinação das técnicas baseadas em características faciais pela regra da soma [45]. O nosso sistema de reconhecimento apresenta um desempenho melhor se comparado ao sistema baseado em ICP, e a combinação das técnicas baseadas em características faciais obteve as melhores taxas de verificação em quase todas as expressões analisadas.

Nossos resultados experimentais mostram um crescimento expressivo nas taxas de verificação quando técnicas baseadas em características faciais são aplicadas no reconhecimento facial. Comparando com os resultados obtidos para a face inteira, utilizando o ICP obtivemos taxas até 46% maiores para expressões intensas (*i.e.* SUR 2), 36% maiores para expressões moderadas (*i.e.* SUR 1) e 5% maior para a expressão neutra. Utilizando o nosso sistema de

Tabela 4.4: Taxas de verificação sob variação de expressões faciais para a face inteira (F), a área circular em torno do nariz (C), a área elíptica em torno do nariz (E), a parte superior da face (U), e a combinação das regiões rígidas (C+E+U).

Método	F	C	E	U	C+E+U
NEU	93.33	97.95	98.35	93.27	98.19
HAP 1	66.94	73.31	83.28	70.89	82.40
HAP 2	23.44	37.30	48.85	32.87	48.28
OPM 1	72.96	91.48	92.59	86.30	95.19
OPM 2	66.58	81.46	91.33	81.46	92.01
FRO 1	74.22	85.16	88.28	84.38	90.63
FRO 2	59.04	78.19	74.47	64.36	84.04
DIS 1	87.21	97.67	97.67	90.70	98.84
DIS 2	42.07	23.17	43.90	37.80	37.20
SUR 1	57.43	89.11	93.32	73.76	93.81
SUR 2	42.53	80.84	80.40	77.68	88.84
SAD 1	75.11	94.05	87.67	82.16	93.61
SAD 2	57.95	79.55	70.83	63.64	78.41
PUF	33.94	42.87	32.58	52.49	53.39
OTH	63.92	72.15	83.86	72.15	83.84

Tabela 4.5: Taxas de verificação sob variações de expressões faciais para as mesmas regiões F, C, E e U da Tabela 4.4, para a abordagem que prioriza os setores rígidos (S) e a combinação das técnicas caseadas em características faciais (C+E+U+S).

Método	F	C	E	U	S	C+E+U+S
NEU	99.40	99.36	99.52	99.32	99.12	99.96
HAP 1	92.52	94.43	95.31	92.08	90.00	98.53
HAP 2	76.89	85.25	89.51	75.41	75.74	94.10
OPM 1	99.26	96.30	96.30	99.26	97.04	97.04
OPM 2	94.73	96.94	99.15	94.05	91.50	99.66
FRO 1	95.31	93.75	92.19	92.19	87.50	100.0
FRO 2	92.02	86.70	87.23	90.96	93.09	90.96
DIS 1	100.0	100.0	97.67	100.0	97.67	100.0
DIS 2	76.83	43.90	57.32	73.17	79.27	74.39
SUR 1	93.07	97.52	99.51	94.55	90.59	100.0
SUR 2	88.23	96.67	97.72	91.21	87.17	98.95
SAD 1	93.39	99.12	100.0	95.15	92.07	100.0
SAD 2	86.36	93.94	94.70	84.10	87.12	96.97
PUF	87.33	88.69	91.40	91.63	86.20	97.51
OTH	88.61	94.94	93.04	90.51	89.87	99.37

reconhecimento, obtivemos taxas até 18% maiores para expressões intensas (*i.e.* HAP 2) e 10% maiores para expressões moderadas (*i.e.* PUF).

4.3 Normalização facial

Um último experimento foi realizado acrescentando-se uma etapa de normalização da pose antes da etapa de reconhecimento para acelerar a convergência dos métodos de registro. Para isto, uma cópia da imagem original invertida no eixo x é criada. Depois esta cópia é alinhada com sua versão original, e a transformação obtida pode ser utilizada para normalizar a imagem original, uma vez que a metade da transformação encontrada (*i.e.* os valores de translação e rotação encontrados são divididos por 2) alinha o plano de simetria da face (*i.e.* plano que divide a face exatamente ao meio) com o plano yz do sistema de coordenadas, como mostrado na Fig. 4.10. Após o alinhamento entre as imagens original e invertida, os pontos na imagem invertida que forem seus próprios pontos correspondentes na imagem original formam a linha de perfil da face, como pode ser visto na Fig. 4.10(b), e o plano formado pelos pontos nessa linha corresponde ao plano de simetria da face. Com este procedimento, a face estará normalizada para translação no eixo x e rotação nos eixos y e z, restando apenas encontrar durante o registro para o reconhecimento os melhores valores para translação nos eixos y e z e rotação no eixo x.

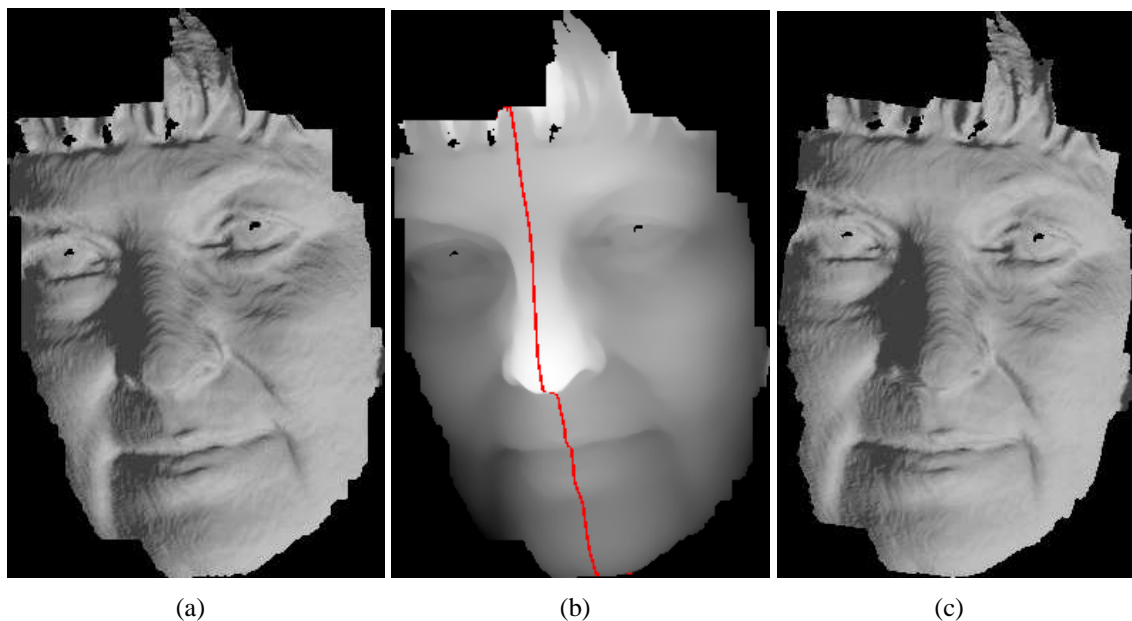


Figura 4.10: Exemplo da aplicação da normalização facial: (a) imagem original, (b) curva de perfil identificada através do processo de normalização, e (c) imagem normalizada.

O nosso sistema de reconhecimento facial foi aplicado em um subconjunto das imagens neutras e sem ruído da base FRGC v2.0 (310 imagens de 62 indivíduos) utilizando a imagem original e depois a sua versão normalizada, com o objetivo de avaliar a diferença de desempenho obtida. Como resultado, o sistema obteve taxas de verificação ligeiramente superiores (*i.e.* aumento de 0.5% considerando FAR de 10^{-4} e de 0.2% considerando FAR de 10^{-3} , como

mostrado na Fig. 4.11) e o tempo de execução foi reduzido em 40% na média, sem que nenhuma alteração fosse realizada no sistema de reconhecimento. Isso foi possível graças à redução do espaço de busca proporcionada pela normalização das imagens de entrada.

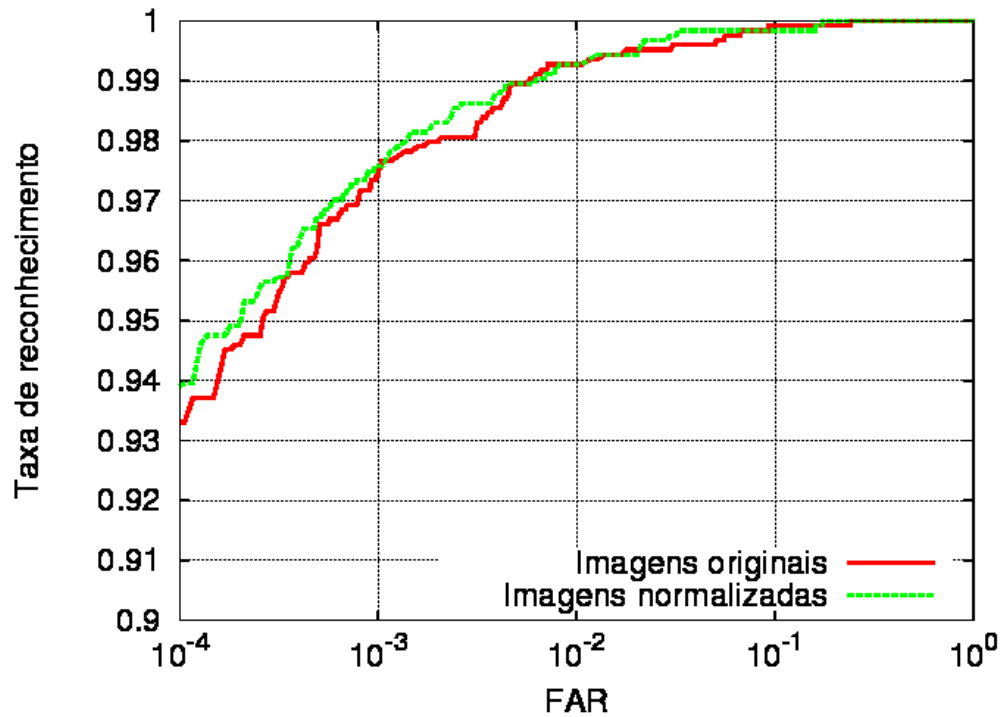


Figura 4.11: Taxas de verificação obtidas para as imagens antes e depois da normalização.

5 Conclusão

Nós apresentamos técnicas automáticas para a segmentação facial e a detecção de características faciais em imagens de profundidade frontalmente posicionadas utilizando somente a informação de profundidade como entrada. Nossa abordagem de segmentação extrai completamente a face combinando detecção de bordas, agrupamento de regiões e análise de forma. Nossa abordagem de detecção de características utiliza classificação de curvatura de superfícies e curvas de relevo das faces segmentadas para detectar a ponta e os cantos do nariz e os cantos internos dos olhos.

A face foi extraída corretamente pela nossa abordagem em 99.3%, 99.7% e 96.1% das imagens das bases FRGC v1.0, FRGC v2.0 e BU-3DFE, respectivamente. Houve um aumento das taxas de verificação em um sistema de reconhecimento baseado em ICP de 2% considerando FAR de 10^{-3} em comparação com duas outras abordagens de segmentação [2,5], e um aumento das taxas de verificação de 0.4% considerando FAR de 10^{-3} quando o nosso sistema de reconhecimento é utilizado.

As características faciais foram corretamente encontradas em 99% das imagens das bases FRGC v1.0 e v2.0, e também 99% das características do nariz e 93% dos cantos dos olhos quando a base BU-3DFE foi utilizada. A ponta do nariz foi a característica facial mais precisa, apresentando os menores valores de erro médio e desvio padrão para todas as bases. A nossa abordagem de detecção de características obteve desempenho equivalente ou melhor quando comparada com três outras abordagens [5–7], e mostrou-se robusta a expressões faciais, como pode ser observado nos experimentos utilizando as bases FRGC v2.0 e BU-3DFE. As características detectadas foram empregadas para extrair três diferentes regiões contendo partes rígidas da face. Estas regiões foram empregadas em um experimento de reconhecimento facial com o objetivo de avaliar a melhoria obtida nas taxas de verificação quando imagens com expressões faciais são utilizadas no reconhecimento. Taxas de verificação até 46% maiores foram obtidas para faces com expressões intensas, 36% maiores para faces com expressão moderada e 5% maior para faces neutras.

Finalmente, nossos resultados mostram os benefícios obtidos por ambas as abordagens de segmentação e detecção de características em um sistema de reconhecimento facial, e o quanto as taxas de verificação podem ser melhoradas. As abordagens apresentadas foram incluídas em nosso sistema de reconhecimento facial 3D [3] baseado na combinação SA+SIM, e nós obtivemos uma taxa de verificação de 99.9% considerando FAR 0 para faces neutras, e 96.5% considerando FAR de 10^{-3} quando todas as imagens da base FRGV v2.0 são comparadas duas a duas (*i.e.* incluindo comparações entre duas imagens não neutras). Estes resultados obtidos de forma completamente automática pelo nosso sistema de reconhecimento facial 3D (*i.e.* *segmentação, detecção de características e comparação*) estão entre os melhores resultados obtidos na literatura [3, 30], e mostram a importância do aprimoramento das etapas de pré-processamento para o reconhecimento.

Como trabalho futuro, pretendemos estender o método para imagens de baixa qualidade, obtidas através de dispositivos de aquisição mais acessíveis (*e.g.* visão estéreo, luz estruturada), adaptando ou substituindo as etapas que dependem de uma informação de profundidade de boa qualidade. O conjunto de características a serem detectadas pode ser ampliado, com o objetivo de utilizar estas características para outras aplicações, como o reconhecimento de expressões faciais. No reconhecimento, podemos avaliar a combinação das abordagens já desenvolvidas com outras abordagens presentes na literatura, como a criação de modelos das deformações causadas por expressões faciais ou a representação da face de uma maneira invariante às expressões faciais.

Referências Bibliográficas

- [1] P. Viola and M. J. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [2] A. Mian, M. Bennamoun, and R. Owens, “An efficient multimodal 2d-3d hybrid approach to automatic face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.
- [3] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. Pamplona Segundo, “3D face recognition using simulated annealing and the surface interpenetration measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 206–219, 2010.
- [4] P. J. Besl, “Surface in range images understanding.” Springer-Verlag, 1988.
- [5] X. Lu and A. K. Jain, “Multimodal facial feature extraction for automatic 3d face recognition,” Department of Computer Science, Michigan State University, Tech. Rep., 2005.
- [6] T.-H. Yu and Y.-S. Moon, “A novel genetic algorithm for 3d facial landmark localization,” in *Proc. of the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.
- [7] M. Romero-Huertas and N. Pears, “3d facial landmark localisation by matching simple descriptors,” in *Proc. of the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.
- [8] O. R. P. Bellon, L. Silva, and C. C. Queirolo, “3D face matching using the surface interpenetration measure,” in *Lecture Notes in Computer Science*. Springer-Verlag, 2005, vol. 3617, pp. 1051–1058.
- [9] X. Lu, A. K. Jain, and D. Colbry, “Matching 2.5D face scans to 3D models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, 2006.
- [10] K. I. Chang, K. W. Bowyer, and P. J. Flynn, “Effects on facial expression in 3D face recognition,” in *Proc. of the SPIE - Biometric Technology for Human Identification*, vol. 5779, 2005, pp. 132–143.
- [11] C. Heshner, A. Srivastava, and G. Erlebacher, “Principal component analysis of range images for facial recognition,” in *Proc. of the International Conference on Imaging Science, Systems and Technology*, 2002.
- [12] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, “Face localization and authentication using color and depth images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 152–168, 2005.

- [13] J. Y. Cartoux, J. T. Lapreste, and M. Richetin, "Face authentication or recognition by profile extraction from range images," in *Proc. of the IEEE Computer Society Workshop on Interpretation of 3D Scenes*, 1989, pp. 194–199.
- [14] G. Pan, Y. Wu, and Z. Wu, "Investigating profile extracted from range data for 3d face recognition," in *Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, 2003, pp. 1396–1399.
- [15] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006.
- [16] X. Lu and A. K. Jain, "Deformation modeling for robust 3d face matching," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1377–1383.
- [17] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-Dimensional face recognition," *International Journal of Computer Vision*, vol. 64, no. 1, pp. 5–30, 2005.
- [18] S. Berreti, A. Del Bimbo, P. Pala, and F. J. S. Mata, "Geodesic distances for 3D-3D and 2D-3D face recognition," in *Proc. of the IEEE International Conference on Multimedia and Expo*, 2007, pp. 1515–1518.
- [19] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, 2007.
- [20] P. J. Besl and R. Jain, "Segmentation through variable-order surface fitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 167–192, 1988.
- [21] A. Colombo, C. Cusano, and R. Schettini, "3d face detection using curvature analysis," *Pattern Recognition*, vol. 39, no. 3, pp. 444–455, 2006.
- [22] A. B. Moreno, A. Sánchez, J. F. Vélez, and F. J. Díaz, "Face recognition using 3d surface-extracted descriptor," in *Proc. of the Irish Machine Vision and Image Processing*, 2003.
- [23] D. Colbry, G. Stockman, and A. K. Jain, "Detection of anchor points for 3d face verification," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005, p. 118.
- [24] T. Faltemier, K. W. Bowyer, and P. J. Flynn, "Rotated profile signatures for robust 3d feature detection," in *Proc. of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–7.
- [25] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3d face recognition," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 585–590.
- [26] P. J. Besl and H. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

- [27] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 211–216.
- [28] T. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using a multi-instance enrollment representation to improve 3d face recognition," in *Proc. of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–6.
- [29] M. Pamplona Segundo, C. C. Queirolo, O. R. P. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," *Proc. of the 14th International Conference on Image Analysis and Processing*, pp. 431–436, 2007.
- [30] M. Pamplona Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics (Aceito)*.
- [31] B. Gökberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3-D face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 38, no. 1, pp. 155–173, 2008.
- [32] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "Bilinear models for 3-D face and facial expression recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 498–511, 2008.
- [33] H. Tang and T. S. Huang, "3D facial expression recognition based on automatically selected features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 23–28.
- [34] —, "3D facial expression recognition based on properties of line segments connecting facial feature points," in *8th IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–6.
- [35] O. R. P. Bellon and L. Silva, "New improvements to range image segmentation by edge detection," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 43–45, 2002.
- [36] A. K. Jain and R. C. Dubes, "Algorithms for clustering data." Prentice-Hall, 1988.
- [37] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967, pp. 281–297.
- [38] R. C. Gonzalez and R. E. Woods, "Digital image processing." Addison-Wesley, 1992.
- [39] L. G. Shapiro and G. C. Stockman, "Computer vision." Prentice-Hall, 2001.
- [40] P. F. Felzenszwalb and D. P. Huttenlocher, "Distance transforms of sampled functions," Cornell Computing and Information Science, Tech. Rep., 2004.
- [41] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Processing-Image Communication*, vol. 12, no. 3, pp. 263–281, 1998.

- [42] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3d face recognition," in *Proc. of the IEEE Workshop FRGC*, 2005.
- [43] P. F. U. Gotardo, O. R. P. Bellon, K. L. Boyer, and L. Silva, "Range image segmentation into planar and quadric surfaces using an improved robust estimator and genetic algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 34, no. 6, pp. 2303–2316, 2004.
- [44] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 63–85, 1997.
- [45] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.